



UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

KAZUMI TOMOYOSE

O *DATA CATALOG VOCABULARY* (DCAT) PARA A PUBLICAÇÃO DE DADOS DE
PESQUISA NOS PRINCÍPIOS *LINKED DATA*

São Carlos/SP
2021

KAZUMI TOMOYOSE

O *DATA CATALOG VOCABULARY* (DCAT) PARA A PUBLICAÇÃO DE DADOS DE PESQUISA NOS PRINCÍPIOS *LINKED DATA*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos (UFSCar), como parte das exigências para a obtenção do título de Mestre em Ciência da Informação.

Área: Conhecimento, Tecnologia e Inovação.

Linha: Tecnologia, Informação e Representação.

Orientadora: Profa. Dra. Ana Carolina Simionato Arakaki.

Financiamento: Processo FAPESP 2018/23655-6.



São Carlos/SP
2021

T661v Tomoyose, Kazumi.
Vocabulários para a publicação de dados de pesquisa nos princípios *Linked Data*: aspectos do *Data Catalog Vocabulary* (DCAT) / Kazumi Tomoyose. — São Carlos, 2021
185 f.

Orientadora: Ana Carolina Simionato Arakaki.
Dissertação (Mestrado em Ciência da Informação)
– Universidade Federal de São Carlos, São Carlos, 2021.

Financiamento: FAPESP (Processo 2018/23655-6)

1. Data Catalog Vocabulary (DCAT). 2. Vocabulários.
3. Dados de pesquisa. 4. Linked Data. I. Título.

CDD 025.3

KAZUMI TOMOYOSE

**O DATA CATALOG VOCABULARY (DCAT) PARA A PUBLICAÇÃO DE DADOS
DE PESQUISA NOS PRINCÍPIOS *LINKED DATA***

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos, como requisito parcial para obtenção do título de Mestrado em Ciência da Informação.

Área de concentração: Conhecimento, Tecnologia e Inovação.

Linha de pesquisa: Tecnologia, Informação e Representação.

BANCA EXAMINADORA

Profa. Dra. Ana Carolina Simionato Arakaki
Universidade Federal de São Carlos (UFSCar)

Prof. Dr. Rogério Aparecido Sá Ramalho
Universidade Federal de São Carlos (UFSCar)

Profa. Dra. Ana Alice Rodrigues Pereira Baptista
Universidade do Minho



UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência da Informação

Folha de Aprovação

Defesa de Dissertação de Mestrado da candidata Kazumi Tomoyose, realizada em 10/03/2021.

Comissão Julgadora:


Profa. Dra. Ana Carolina Simionato Arakaki (UFSCar)

Profa. Dra. Ana Alice Rodrigues Pereira Baptista (UMinho)

Prof. Dr. Rogério Aparecido Sá Ramalho (UFSCar)

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

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Informação.

AGRADECIMENTOS

Aos membros da banca, Profa. Dra. Ana Alice Rodrigues Pereira Baptista, Prof. Dr. Rogério Aparecido Sá Ramalho, Prof. Dr. José Eduardo Santarém Segundo e Prof. Dr. Fabiano Ferreira de Castro, pela dedicação de seu tempo e leitura da dissertação.

À minha orientadora Profa. Dra. Ana Carolina Simionato Arakaki pelo acompanhamento e auxílio no decorrer da pesquisa.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo apoio financeiro para o desenvolvimento dessa pesquisa (Processo FAPESP 2018/23655-6).

RESUMO

Com o desenvolvimento das tecnologias semânticas, cresce a importância e o reconhecimento dos dados como fontes informacionais. Nesse contexto, os dados gerados a partir de pesquisas são cada vez mais valorizados, uma vez que sua publicação permite a fundamentação dos resultados obtidos com os trabalhos científicos, além de auxiliarem na produção de novos conhecimentos de forma sustentável. Assim, é observada a necessidade em organizar esses dados de pesquisa, de forma a otimizar a recuperação deles, sendo que poderiam ser utilizados, para tal finalidade, os vocabulários. Os vocabulários são compreendidos como conjuntos de termos que padronizam a descrição de recursos, sendo o foco da presente pesquisa aqueles voltados à padronização dos elementos de metadados, pensando-se na organização e recuperação da informação. Ao considerar o ambiente da *Web*, a utilização dos princípios *Linked Data* melhora a descoberta de novos dados, sendo que o *Data Catalog Vocabulary* (DCAT) figura como um vocabulário voltado à descrição de catálogos e seus conjuntos de dados, buscando favorecer a interoperabilidade na *Web*. O objetivo da pesquisa consiste em analisar as características e os fundamentos do DCAT e dos vocabulários adotados por ele para a publicação de dados de pesquisa nos princípios *Linked Data*, de modo a identificar a discussão sobre a temática na literatura e verificar a utilização do mesmo. Por meio de pesquisa exploratória, descritiva e teórica, foi possível sistematizar e analisar as características dos vocabulários, identificando-se que o DCAT permite que os catálogos e os conjuntos de dados sejam padronizados, diminuindo questões de ambiguidade e proporcionando a descoberta de dados e de serviços de dados, uma vez que dispõe termos específicos para tal finalidade. Desse modo, alinhado também aos princípios *Linked Data*, o uso do DCAT permite que sejam efetivados os benefícios do compartilhamento dos dados de pesquisa. Foi realizada também a análise de plataformas do re3data a fim de identificar a atual utilização do vocabulário por elas, verificando-se baixa adoção do DCAT. Entretanto, os resultados indicam a preocupação dessas plataformas com questões de encontrabilidade, acessibilidade, interoperabilidade e reutilização, indo de encontro com os princípios *Linked Data* e os princípios FAIR. A pesquisa revelou que existem poucas discussões sobre o DCAT na literatura, em especial no âmbito da Ciência da Informação. Dessa forma, identifica-se a necessidade de melhor explorar as possibilidades de representação da informação no contexto dos dados de pesquisa. Além disso, maiores discussões que alinhem o conhecimento da área às ferramentas e aos conceitos semânticos aplicados à *Web* potencializam os benefícios de encontrabilidade, utilização e recuperação de informações, além de expandir e aprimorar os conhecimentos e as possibilidades de atuação dos profissionais da informação.

Palavras-chave: Data Catalog Vocabulary (DCAT). Vocabulários. Dados de pesquisa. Linked Data.

ABSTRACT

With the development of semantic technologies, the importance and the recognition of data as information sources grow. In this context, the data generated from researches are increasingly valued, since their publication allows the foundation of the results obtained with the scientific works, also helping in the production of new knowledge in a sustainable way. Thus, there is a need to organize these research data, so as to optimize their retrieval. For this purpose, vocabularies could be used. From the standardization of data representation, vocabularies allow data to be organized and retrieved efficiently. When considering the Web environment, the discovery of new data is also improved, using Linked Data principles, and the Data Catalog Vocabulary (DCAT) appears as a vocabulary aimed at describing catalogs and its datasets, seeking to favor their interoperability. The objective of the research is to analyze the characteristics and fundamentals of the Data Catalog Vocabulary (DCAT) and the vocabularies adopted by it, for the publication of research data in the Linked Data principles, in order to identify the discussion on the subject in the literature and verify the use of it. Through exploratory, descriptive and theoretical research, it was possible to systematize and analyze the characteristics of the vocabularies, identifying that DCAT allows catalogs and datasets to be standardized, reducing ambiguity issues and providing the discovery of data and data services, since it has specific terms for this purpose. Thus, in line with Linked Data principles, the use of DCAT allows the benefits of sharing research data to be accomplished. An analysis of re3data platforms was also carried out in order to identify the current use of vocabulary by them, where we verified low adoption of DCAT. However, the results indicate the concern of these platforms with issues of findability, accessibility, interoperability and reuse, going against the Linked Data principles and the FAIR principles. The research revealed that there are few discussions about DCAT in the literature, especially in the scope of Information Science. Thus, we point to the need to better explore the possibilities of representing information in the context of research data. In addition, further discussions that align the knowledge of the area with the semantic tools and concepts applied to the Web enhance the benefits of finding, using and retrieving information, in addition to expanding and improving the knowledge and possibilities of information professionals' performance.

Keywords: Data Catalog Vocabulary (DCAT). Vocabularies. Research data. Linked Data.

LISTA DE FIGURAS

Figura 1 – <i>Digital Curation Centre (DCC) Curation Lifecycle Model</i>	40
Figura 2 – Ciclo de vida de dados do <i>Research360</i>	43
Figura 3 – Bolo de noiva da <i>Web Semântica</i> , por Berners-Lee (2000)	51
Figura 4 – Pilha de tecnologias da <i>Web Semântica</i> , por Nowack (2009).....	52
Figura 5 – Bolo de noiva de tecnologias da <i>Web Semântica</i> ajustado, por Idehen (2017).....	53
Figura 6 – <i>Linked Data</i> e <i>Linked Open Data</i>	59
Figura 7 – <i>Linked Open Data Cloud (LOD Cloud)</i> de 26 de janeiro de 2021	61
Figura 8 – Ciclo de vida <i>Linked Data</i>	62
Figura 9 – Exemplo de uso do <i>XML Schema</i>	69
Figura 10 – Exemplo de uso de classe do <i>SKOS</i>	73
Figura 11 – Visão geral do <i>Data Catalog Vocabulary (DCAT)</i>	92
Figura 12 – Exemplo de uso do <i>FOAF Vocabulary</i>	106
Figura 13 – Exemplo de uso do <i>vCard Ontology</i>	113
Figura 14 – Exemplo de uso de metadados gerais do <i>Vocabulary of Interlinked Datasets (VOID)</i>	121
Figura 15 – Exemplo de uso de metadados de acesso do <i>Vocabulary of Interlinked Datasets (VOID)</i>	121
Figura 16 – Exemplo de uso de metadados estruturais do <i>Vocabulary of Interlinked Datasets (VOID)</i>	122
Figura 17 – Exemplo de uso de metadados para descrição de <i>linksets</i> do <i>Vocabulary of Interlinked Datasets (VOID)</i>	122
Figura 18 – Diagrama das relações entre as classes e propriedades do <i>Vocabulary of Interlinked Datasets (VOID)</i>	126
Figura 19 – Busca pelo <i>Data Catalog Vocabulary</i> , realizada no <i>re3data</i>	127
Figura 20 – Registro do repositório <i>Ag Data Commons</i>	129
Figura 21 – Metadados em formato <i>RDF</i> de registro do repositório <i>Ag Data Commons</i>	129
Figura 22 – Registro do catálogo <i>Arctic Permafrost Geospatial Centre</i>	130
Figura 23 – Metadados em formato <i>RDF/XML</i> de registro do catálogo <i>Arctic Permafrost Geospatial Centre</i>	131

Figura 24 – Registro do repositório <i>Biological and Schemical Oceanography Data Management Office</i>	132
Figura 25 – Metadados em formato RDF/XML de registro do repositório <i>Biological and Chemical Oceanography Data Management Office</i>	132
Figura 26 – Registro do repositório <i>COEMS Open Data</i>	133
Figura 27 – Metadados em formato RDF/XML de registro do repositório <i>COEMS Open Data</i>	134
Figura 28 – Registro de <i>Data.gov.au</i>	135
Figura 29 – Metadados em formato RDF/XML de registro do <i>Data.gov.au</i>	135
Figura 30 – Elemento ‘ <i>Research Dataset</i> ’ do Metax	136
Figura 31 – Registro do <i>European Data Portal</i>	137
Figura 32 – Metadados em formato JSON-LD do <i>European Data Portal</i>	138
Figura 33 – Busca em SPARQL no <i>European Union Open Data Portal</i>	139
Figura 34 – Registro do <i>GovData</i>	140
Figura 35 – Metadados em formato RDF/XML do <i>GovData</i>	140
Figura 36 – Registro do <i>ICOS Carbon Portal</i>	141
Figura 37 – Metadados em formato RDF/XML do <i>ICOS Carbon Portal</i>	142
Figura 38 – Metadados em formato RDF/XML de registro do <i>JRC Data Catalogue</i>	143
Figura 39 – Registro de <i>Norwegian Polar Data Centre</i>	144
Figura 40 – Metadados em formato JSON-LD de registro do <i>Norwegian Polar Data Centre</i>	144
Figura 41 – Seção de dados do <i>Open Data Kingston</i>	146
Figura 42 – Metadados em formato RDF/XML do catálogo do <i>Open Data Kingston</i>	146
Figura 43 – Registro do <i>Open Government Canada – Open Data</i>	147
Figura 44 – Metadados no formato RDF/XML do <i>Open Government Canada – Open Data</i>	148
Figura 45 – Registro de <i>RDPCIDAT</i>	149
Figura 46 – Metadados no formato RDF/XML de registro do <i>RDPCIDAT</i>	149
Figura 47 – Registro do catálogo do <i>Répertoire International des Sources Musicales</i>	150
Figura 48 – Metadados no formato RDF/XML de registro do <i>Répertoire International des Sources Musicales</i>	151

Figura 49 – Busca por SPARQL <i>endpoint</i> do <i>Répertoire International des Sources Musicales</i>	152
Figura 50 – Metadados no formato RDF do catálogo do <i>Répertoire International des Sources Musicales</i>	153
Figura 51 – Metadados em formato XHTML+RDFa de registro do <i>Repositorio de datos del Ministerio de Educación del Perú</i>	154

LISTA DE QUADROS

Quadro 1 – Categorias de metadados na literatura.....	30
Quadro 2 – Cobertura de políticas para compartilhamento de dados de pesquisa de instituições financiadoras nacionais	49
Quadro 3 – Sistema de classificação por estrelas para <i>Linked Open Data</i>	59
Quadro 4 – Classes principais do <i>Resource Description Framework Schema</i> (RDF Schema).....	65
Quadro 5 – Classes adicionais do <i>Resource Description Framework Schema</i> (RDF Schema).....	65
Quadro 6 – Propriedades principais do <i>Resource Description Framework Schema</i> (RDF Schema)	66
Quadro 7 – Propriedades adicionais do <i>Resource Description Framework Schema</i> (RDF Schema)	67
Quadro 8 – Elementos do XML Schema	68
Quadro 9 – Atributos do XML Schema.....	69
Quadro 10 – Classes do SKOS.....	71
Quadro 11 – Propriedades do SKOS	71
Quadro 12 – Princípios FAIR.....	74
Quadro 13 – Melhores práticas para publicação de dados na Web	77
Quadro 14 – Tipos de padrões de dados	83
Quadro 15 – Classes do <i>Data Catalog Vocabulary</i> (DCAT).....	85
Quadro 16 – Propriedades do <i>Data Catalog Vocabulary</i> (DCAT)	86
Quadro 17 – Exemplo de uso do <i>Data Catalog Vocabulary</i> (DCAT) em catálogo de dados	93
Quadro 18 – Exemplo de uso do <i>Data Catalog Vocabulary</i> (DCAT), com indicação de proveniência	93
Quadro 19 – Exemplo de uso do <i>Data Catalog Vocabulary</i> (DCAT), com ligação entre conjunto de dados e publicação.....	94
Quadro 20 – Exemplo de uso do <i>Data Catalog Vocabulary</i> (DCAT) para descrição de serviços de dados	95
Quadro 21 – Exemplo de uso do <i>Data Catalog Vocabulary</i> (DCAT) para descrição de distribuições comprimidas e em pacote	95
Quadro 22 – Tipos de termos de metadados mantidos pela DCMI.....	99

Quadro 23 – Exemplo de uso do DCMI <i>Metadata Terms</i>	100
Quadro 24 – Classes do <i>Friend of a Friend Vocabulary</i> (FOAF <i>Vocabulary</i>).....	102
Quadro 25 – Propriedades do <i>Friend of a Friend Vocabulary</i> (FOAF <i>Vocabulary</i>) .	103
Quadro 26 – Classes do <i>vCard Ontology</i>	107
Quadro 27 – Propriedades de objeto do <i>vCard Ontology</i>	110
Quadro 28 – Propriedades de dados do <i>vCard Ontology</i>	111
Quadro 29 – Classes de ponto de partida do <i>PROV Ontology</i> (PROV-O).....	114
Quadro 30 – Propriedades de ponto de partida do <i>PROV Ontology</i> (PROV-O).....	115
Quadro 31 – Classes expandidas do <i>PROV Ontology</i> (PROV-O).....	116
Quadro 32 – Propriedades expandidas do <i>PROV Ontology</i> (PROV-O).....	116
Quadro 33 – Classes qualificadas do <i>PROV Ontology</i> (PROV-O).....	117
Quadro 34 – Propriedades qualificadas do <i>PROV Ontology</i> (PROV-O).....	118
Quadro 35 – Classes do <i>Vocabulary of Interlinked Datasets</i> (VOID).....	123
Quadro 36 – Propriedades do <i>Vocabulary of Interlinked Datasets</i> (VOID).....	123
Quadro 37 – Resultado da análise de utilização do <i>Data Catalog Vocabulary</i> (DCAT) nas plataformas registradas no re3data.....	155

LISTA DE ABREVIATURAS E SIGLAS

AKSW	<i>Agile Knowledge Engineering and Semantic Web</i>
DC Terms	<i>Dublin Core Metadata Initiative Metadata Terms</i>
DCAM	<i>DCMI Abstract Model</i>
DCAP	<i>Dublin Core Application Profile</i>
DCAT	<i>Data Catalog Vocabulary</i>
DCC	<i>Digital Curation Centre</i>
DCMES	<i>Dublin Core Metadata Element Set</i>
DCMI	<i>Dublin Core Metadata Initiative</i>
DOI	<i>Digital Object Identifier</i>
EU ODP	<i>European Union Open Data Portal</i>
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FAIR	<i>Findable, accessible, interoperable and reusable</i>
FOAF	<i>Friend of a Friend</i>
HTTP	<i>Hypertext Transfer Protocol</i>
ICOS	<i>Integrated Carbon Observation System</i>
IRI	<i>International Resource Identifier</i>
JRC	<i>Joint Research Centre</i>
LOD	<i>Linked Open Data</i>
MP	Melhor Prática
OWL2	<i>OWL2 Web Ontology Language</i>
PGD	Plano de Gestão de Dados
PROV- CONSTRAINTS	Constraints of the PROV Data Model
PROV-DM	<i>PROV Data Model</i>
PROV-N	<i>Provenance Notation</i>
PROV-O	<i>PROV Ontology</i>
RDF	<i>Resource Description Framework</i>
RDFa	<i>Resource Description Framework in Attributes Core</i>

RDFS	<i>Resource Description Framework Schema</i>
RDPCIDAT	<i>Research Data Repository – Research Department Plasmas with Complex Interactions</i>
re3data	<i>Registry of Research Data Repositories</i>
RISM	<i>Répertoire International des Sources Musicales</i>
SKOS	<i>Simple Knowledge Organization System</i>
SPARQL	<i>SPARQL Protocol and RDF Query Language</i>
TIC	<i>Tecnologias de Informação e Comunicação</i>
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
URN	<i>Uniform Resource Name</i>
VoID	<i>Vocabulary of Interlinked Datasets</i>
W3C	<i>World Wide Web Consortium</i>
WoDQA	<i>Web of Data Query Analyzer</i>
XHTML	<i>eXtensible Hypertext Markup Language</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Justificativa	14
1.2 Objetivos	15
1.3 Procedimentos metodológicos	16
1.4 Plano de trabalho	17
1.5 Estrutura da dissertação	18
2 A REPRESENTAÇÃO DA INFORMAÇÃO NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO	20
3 OS DADOS DE PESQUISA	32
4 PRINCÍPIOS <i>LINKED DATA</i> E O COMPARTILHAMENTO DE DADOS.....	50
4.1 Tecnologias semânticas.....	64
4.1.1 <i>RDF Schema</i>	64
4.1.2 <i>XML Schema</i>	68
4.1.3 <i>Simple Knowledge Organization System (SKOS)</i>	70
4.2 Princípios FAIR	73
4.3 Melhores práticas para publicação de dados na <i>Web</i>	76
5 <i>DATA CATALOG VOCABULARY (DCAT)</i>	82
5.1 <i>Dublin Core Metadata Initiative Metadata Terms (DC Terms)</i>	97
5.2 <i>Friend of a Friend Vocabulary (FOAF Vocabulary)</i>	102
5.3 <i>vCard Ontology</i>	107
5.4 <i>PROV Ontology (PROV-O)</i>	113
5.5 <i>Vocabulary of Interlinked Datasets (VOID)</i>	120
6 UTILIZAÇÃO DO <i>DATA CATALOG VOCABULARY (DCAT)</i>	127
6.1 <i>Ag Data Commons</i>	128
6.2 <i>Arctic Permafrost Geospatial Centre</i>	129
6.3 <i>Biological and Chemical Oceanography Data Management Office</i>	131

6.4 COEMS <i>Open Data</i>	133
6.5 Data.gov.au.....	134
6.6 <i>Etsin Research Data Finder</i>	135
6.7 <i>European Data Portal</i>	137
6.8 <i>European Union Open Data Portal</i>	138
6.9 <i>Fairdata IDA Research Data Storage Service</i>	139
6.10 GovData.....	139
6.11 <i>ICOS Carbon Portal</i>	141
6.12 <i>JRC Data Catalogue</i>	142
6.13 <i>Norwegian Polar Data Centre</i>	143
6.14 <i>Open Data by Socrata</i>	145
6.15 <i>Open Data Kingston</i>	145
6.16 <i>Open Government Canada - Open Data</i>	147
6.17 RDPCIDAT.....	148
6.18 <i>Répertoire International des Sources Musicales (RISM)</i>	150
6.19 <i>Repositorio de datos del Ministerio de Educación del Perú</i>	153
6.20 Discussão sobre as análises das plataformas.....	155
7 CONSIDERAÇÕES FINAIS.....	157
REFERÊNCIAS.....	160

1 INTRODUÇÃO

O desenvolvimento das tecnologias semânticas disponíveis gera um contínuo impacto em diferentes esferas da sociedade, principalmente no meio informacional. Devido ao constante aumento no número de informações geradas, cresce, conjuntamente, a necessidade dos usuários de identificar fontes de informação fidedignas. Nesse cenário, as pesquisas científicas se apresentam como uma fonte capaz de fornecer informações com maior veracidade e segurança.

A pesquisa consiste na investigação científica em um determinado campo do conhecimento para responder a um determinado problema, adotando-se um conjunto de procedimentos metodológicos que delimitam a pesquisa. Por meio da comunicação científica, entre periódicos ou eventos, os resultados são divulgados com o objetivo de compartilhar as descobertas e constatações obtidas (PITTA; CASTRO, 2006). A disponibilização é feita, majoritariamente, de forma *online*, acompanhando o avanço do ambiente *Web*, facilitando o acesso às pesquisas.

Além da divulgação das próprias pesquisas, mostra-se necessária a publicação da pesquisa de forma ampliada, isto é, deve ocorrer a publicação dos dados utilizados para gerá-las, uma vez que também são fontes de informação pertinentes que permitem um melhor entendimento dos processos que culminaram no desenvolvimento da pesquisa, bem como sustentam novas produções científicas (GRAY *et al.*, 2002).

A associação dos princípios *Linked Data* à publicação de dados de pesquisa pode ampliar as possibilidades de uso e reuso dos mesmos, potencializando o desenvolvimento da Ciência, a partir do estabelecimento de ligações semânticas entre dados relacionados, otimizando a sua descoberta e recuperação (BERNERS-LEE, 2006).

Nesse contexto, identifica-se a necessidade da utilização de formas padronizadas de se organizar os dados publicados, levantando-se, aqui, os vocabulários. Eles permitem que os metadados que constituem os dados sejam uniformizados, garantindo consistência em sua organização e recuperação, bem como nas relações *Linked Data*.

Ressalta-se que o termo 'vocabulário' pode ser compreendido de formas diferentes para as diversas áreas do conhecimento (BAKER; VANDENBUSSCHE; VATANT; 2013), sendo que, no contexto da presente pesquisa, vocabulário é compreendido como um conjunto de termos que visa a descrição de algum recurso

(WORLD WIDE WEB CONSORTIUM, c2015c). Assim, tem-se os vocabulários voltados à padronização dos elementos de metadados e os que buscam uniformizar os valores, sendo os primeiros os objetos de estudo dessa pesquisa.

Dessa forma, buscou-se analisar o *Data Catalog Vocabulary* (DCAT) no contexto da publicação de dados de pesquisa, assim como os principais vocabulários que ele adota, para a organização e padronização desses dados, alinhados aos princípios *Linked Data*, a fim de compreender como os vocabulários estão envolvidos na recuperação dos dados de pesquisa, bem como os possíveis benefícios proporcionados pelos mesmos.

1.1 Justificativa

A publicação dos dados utilizados no decorrer de uma pesquisa tem se tornado cada vez mais requisitado, uma vez que se trata de importantes insumos cuja consulta pode auxiliar na assimilação e averiguação da pesquisa, além de servirem de base para o desenvolvimento de novos estudos (COSTELLO, 2009). Assim, encontra-se em crescimento o número de entidades, em diferentes áreas do conhecimento, que reconhecem e demandam a publicação de dados como atividade ligada à divulgação científica (COSTELLO, 2009). Como exemplo, cita-se a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) que passou, em alguns casos, a cobrar o compartilhamento de dados oriundos de pesquisas fomentadas pela instituição (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, [2018?]).

O tema proposto justifica-se, portanto, a partir de sua crescente demanda e de escassas investigações a respeito na área de Ciência da Informação, especialmente em âmbito nacional, necessitando-se, também, de formas de organização dos dados de pesquisa. A presente pesquisa contribui no desenvolvimento de novos conhecimentos acerca da publicação de dados de pesquisa no contexto da Ciência da Informação, considerando-se o cenário das tecnologias semânticas ao trazer juntamente em foco os princípios *Linked Data*.

Levanta-se que com a aplicação de vocabulários para a publicação de dados de pesquisa otimiza-se a descoberta e a recuperação da informação, de forma que os usuários possam obter informações com maior eficiência. Do ponto de vista social, a otimização no tratamento dos dados de pesquisa para a sua publicação

proporciona benefícios ao acesso à informação para a sociedade em geral, ao melhorar o acesso a informações verídicas e seguras.

Dessa maneira, aprimora-se a qualidade das plataformas que conservam e disponibilizam os dados publicados, melhorando, ainda, questões como a visibilidade dos dados. Um fator benéfico às instituições é a sustentabilidade que a utilização de vocabulários fornece por meio da padronização, evitando-se o retrabalho.

Aponta-se que a temática se encontra de acordo com o tema do Grupo de Pesquisa Dados e Metadados (GPDM) da UFSCar, do qual a mestranda participa, ao passo que explora os dados científicos como recursos informacionais e a sua organização e representação.

Além disso, o presente projeto permite dar continuidade à temática abordada durante o período de Iniciação Científica (2016/20665-5), de representação da informação na *Web*, realizando um maior aprofundamento nas iniciativas, princípios, instrumentos e tecnologias semânticas aplicados ao campo da Ciência da Informação, alinhados à publicação de dados de pesquisa, de forma a proporcionar novos olhares à área, bem como evidenciar a contribuição que a Ciência da Informação pode trazer para a própria *Web*.

1.2 Objetivos

O objetivo geral da pesquisa compreende em analisar as características e os fundamentos do DCAT para a publicação de dados de pesquisa nos princípios *Linked Data*, enquanto os objetivos específicos consistem em:

- Realizar maior aprofundamento teórico sobre: vocabulários direcionados a dados, dados de pesquisa e princípios *Linked Data*;
- Descrever os principais vocabulários utilizados pelo DCAT: *Dublin Core Metadata Initiative Metadata Terms (DC Terms)*, *Friend of a Friend Vocabulary (FOAF Vocabulary)*, *vCard Ontology*, *PROV Ontology (PROV-O)* e *Vocabulary of Interlinked Datasets (VOID)*. Investigar, também, o *RDF Schema*, o *XML Schema* e o *Simple Knowledge Organization System (SKOS)*;
- Verificar a utilização do DCAT para a publicação de dados de pesquisa em plataformas digitais, recorrendo-se ao re3data como base de consulta.

1.3 Procedimentos metodológicos

A pesquisa se caracteriza como qualitativa e exploratória, de natureza teórica, abordando os tópicos: publicação de dados de pesquisa, vocabulários voltados à publicação de dados e os princípios *Linked Data*.

O presente trabalho configura-se, ainda, como pesquisa bibliográfica, para a qual foram utilizadas fontes bibliográficas primárias e secundárias a fim de contextualizar teoricamente as temáticas. De acordo com Gil (2002, p. 44), a pesquisa bibliográfica é aquela "[...] desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos.". Trata-se de uma forma de reunir os materiais sobre determinado tema, que se encontravam dispersos, para aumentar a 'área de abrangência da pesquisa' (MINUSI *et al.*, 2018, n.p.). Dessa forma, facilita-se o acesso às informações, uma vez que "A principal vantagem da pesquisa bibliográfica reside no fato de permitir ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que poderia pesquisar diretamente." (GIL, 2002, p. 44).

O material utilizado consistiu em publicações nacionais e internacionais, nos idiomas português, inglês e espanhol, publicados a partir do ano 2000, momento em que se identifica o crescimento de divulgações sobre a publicação de dados e, a partir de 2006, sobre os princípios *Linked Data*. Durante a pesquisa foram utilizadas as bases de dados: Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), Portal de Periódicos da CAPES, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), *Scientific Electronic Library Online* (SciELO), *Scopus* e *Web of Science*. Foram acompanhados e consultados, ainda, os relatórios e notas do W3C, assim como os anais dos eventos: *Association for Information Science and Technology (ASIS&T) Annual Meeting, Dublin Core Metadata Initiative (DCMI) Annual Conference* e Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB). As consultas foram realizadas no período de março de 2019 a dezembro de 2020.

Para a busca foram empregados determinados termos e expressões, a fim de obter materiais que correspondem ao tema da pesquisa, sendo: 'dados de pesquisa', 'dados ligados', 'vocabulários' e 'publicação de dados', bem como suas respectivas traduções ao inglês, 'research data', 'Linked Data', 'vocabularies' e 'data publishing', e ao espanhol, 'datos de investigación', 'datos vinculados', 'vocabularios' e 'publicación de datos'. Observa-se que as expressões equivalentes a 'dados de

pesquisa' também foram buscadas, sendo elas: 'dados científicos', '*scientific data*' e '*datos científicos*'. Como estratégias de busca foram utilizados os *booleanos* 'AND' e 'OR', sendo este empregado em casos de terminologias sinônimas.

Para satisfazer o objetivo de 'verificar a utilização de vocabulários para a publicação de dados de pesquisa em plataformas digitais' foi realizado o levantamento e a análise de repositórios e catálogos que fazem uso do DCAT. Utilizou-se como referência as plataformas registradas no re3data¹, uma vez que possui caráter internacional, permitindo, possivelmente, um panorama mais abrangente da adoção do vocabulário. A averiguação da utilização ou não do DCAT por cada uma das plataformas se deu a partir da análise dos metadados dos registros, sendo que os registros foram escolhidos de forma aleatória. Quando os registros não estavam disponíveis para *download*, foi realizada tentativa de contato por *e-mail* com os administradores de cada uma das plataformas. Essa análise foi realizada durante o período de novembro de 2020 a fevereiro de 2021.

1.4 Plano de trabalho

Com base nos procedimentos metodológicos adotados para a realização da pesquisa, sua execução dividiu-se nas seguintes etapas:

1ª etapa – Levantamento bibliográfico e seleção do material obtido: investigou-se as informações relevantes ao tema da pesquisa com o objetivo de desenvolver a sustentação teórica necessária ao estudo, a partir da busca bibliográfica em nível nacional e internacional, considerando-se materiais impressos e digitais. Para a realização da busca foram utilizados os termos e palavras-chave expostos em '3 – Procedimentos metodológicos'. Para a seleção dos materiais foram consideradas sua relevância e atualidade, assim como a aderência ao tema. Essa etapa abrangeu todos os objetivos do trabalho.

2ª etapa – Interpretação, análise e sistematização das informações: a fim de oportunizar o desenvolvimento da base teórica, provendo os insumos necessários para a discussão dos diferentes posicionamentos identificados na literatura sobre o tema em questão, abrangendo o primeiro e o segundo objetivo dessa pesquisa.

¹ Disponível em: <https://www.re3data.org/>.

3ª etapa – Análise e estabelecimento das características fundamentais extraídas da literatura: para estabelecimento da solução do problema de pesquisa, verificando as características principais encontradas na literatura, abrangendo, assim, o primeiro e o segundo objetivo desta pesquisa.

4ª etapa – Sistematização do estudo exploratório: para análise dos vocabulários e os relacionamentos com a publicação de dados de pesquisa, abrangendo, assim, o segundo e o terceiro objetivo desta pesquisa.

5ª etapa – Elaboração e redação final da pesquisa: desenvolvimento do documento textual com as considerações finais da pesquisa para divulgação à comunidade científica dos resultados obtidos com o desenvolvimento do estudo em questão, abrangendo todos os objetivos desta pesquisa.

6ª etapa – Divulgação da pesquisa: essa etapa não se constitui unicamente do momento final da pesquisa, mas sim, de um processo construtivo e contínuo conjunto do orientador e aluno no decorrer da investigação.

1.5 Estrutura da dissertação

A dissertação apresenta, além da presente seção correspondente à introdução, com justificativa, objetivos, procedimentos metodológicos e plano de trabalho, a seguinte estruturação:

A **Seção 2 – A REPRESENTAÇÃO DA INFORMAÇÃO NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO** apresenta uma contextualização da área na ciência e discorre sobre a representação da informação, abordando os metadados.

A **Seção 3 – OS DADOS DE PESQUISA** define os dados de pesquisa e suas características, apresentando, também, o Plano de Gestão de Dados e o ciclo de vida dos dados.

A **Seção 4 – PRINCÍPIOS LINKED DATA E O COMPARTILHAMENTO DE DADOS** aborda a definição de *Linked Data*, bem como as tecnologias envolvidas – *RDF Schema*, *XML Schema* e *SKOS*. Além disso, discorre sobre os princípios *FAIR* e as melhores práticas para publicação de dados na *Web*.

A **Seção 5 – DATA CATALOG VOCABULARY (DCAT)** apresenta a investigação realizada quanto ao *DCAT*, assim como dos vocabulários *DC Terms*, *FOAF Vocabulary*, *VOID*, *vCard Ontology* e *PROV-O*.

A **Seção 6 – UTILIZAÇÃO DO DATA CATALOG VOCABULARY (DCAT)** apresenta a análise das plataformas registradas no re3data que utilizam o DCAT e a discussão dessa análise.

A **Seção 7 – CONSIDERAÇÕES FINAIS** apresenta as observações e reflexões finais com base no material levantado.

2 A REPRESENTAÇÃO DA INFORMAÇÃO NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO

O desenvolvimento da ciência e das diferentes áreas do conhecimento é imprescindível quando se busca compreender o seu funcionamento. Ao analisar as suas partes é possível identificar suas características e melhor defini-la, tanto em sua forma geral – objetivo e escopo – quanto em seus focos específicos – temáticas.

A ciência configura-se como parte inerente do desenvolvimento da sociedade, sendo que a forma como é desenvolvida e seus resultados são utilizados para melhorar diferentes aspectos de seus indivíduos. Nesse sentido, Latour (2000) sinaliza a importância de se compreender como a ciência ocorre. Segundo o autor, existem dois momentos: a ciência pronta, em que os desenvolvimentos científicos e tecnológicos estão prontos e funcionam; e a ciência em construção, em que se busca entender os detalhes que constituem a ciência. Latour (2000) defende o estudo da ciência de acordo com a última perspectiva, de uma ciência em construção, buscando estudar um problema em uma retrospectiva temporal, visando a compreensão da ciência, em vez da aceitação de uma ciência pronta.

Essa concepção de ciência em construção apresenta similaridades com o que Kuhn (2006) coloca como ‘revoluções científicas’, que trazem a ruptura da ciência determinada como normal, padrão, introduzindo os questionamentos e razões do que a ciência é, e o que é, em uma análise minuciosa que visa construir uma visão capaz de responder aos problemas que a ciência de norma não conseguira fazer. Trata-se da ruptura de um paradigma² vigente e o surgimento de outro, de uma nova forma de se olhar para o fazer científico, complementando as lacunas deixadas pelo paradigma anterior.

Latour (2000) aborda que a ciência é uma construção social, construída pela sociedade e para ela. Assim, um produto apresenta, além de peças e tecnologias dispostas de determinada forma, um reflexo da instituição responsável por seu desenvolvimento, o sistema adotado, os protocolos, demonstrando que “Contexto e conteúdo se confundem” (LATOUR, 2000, p. 18). A visão da ciência em sua complexidade, envolvendo os aspectos sociais, é também abordada por Bourdieu (1976), que coloca o ambiente científico como campo social, sendo um ambiente

² “[...] o paradigma é um modelo que nos permite ver uma coisa em analogia a outra.” (CAPURRO, 2003).

marcado por luta, em que os pesquisadores buscam por prestígio científico para crescerem e se posicionarem nesse âmbito.

Para tratar dos diferentes tipos de conhecimento temos diferentes ciências ou disciplinas. Essas se estabelecem como ‘ciência moderna’ ou ‘ciência pós-moderna’. As ciências modernas possuem um foco maior na funcionalidade, “[...] impulsionada pela busca pela compreensão plena de como o mundo funciona [...]” (WERSIG, 1992, p. 229, tradução nossa), enquanto as ciências pós-modernas visam solucionar os problemas deixados pelas ciências modernas e suas tecnologias, sendo “[...] um conhecimento sobre as condições de possibilidades o que, no mínimo, gera complicadores metodológicos.”, devido ao seu caráter “compreensivo e interpretativo” (TÁLAMO; SMIT, 2007, p. 38).

Desse modo, a Ciência da Informação (CI) nasce como uma ciência pós-moderna, a partir da necessidade de uma forma de lidar com o crescente número de informações que surgem a partir da explosão informacional após a Segunda Guerra Mundial (NHACUONGUE; FERNEDA, 2015). Borko (1968, p. 5, tradução nossa) define a CI como:

[...] uma ciência interdisciplinar que investiga as propriedades e comportamento da informação, as forças que governam o fluxo e uso da informação, e as técnicas, tanto manuais quanto mecânicas, do processamento da informação para armazenamento ótimo, recuperação e disseminação.

Nhacuongue e Ferneda (2015, p. 9) trazem a definição de que “A Ciência da Informação é definida a partir dos problemas que se predispõe a resolver e a sua estrutura está baseada no vínculo permanente entrelaçado entre a pesquisa empírica e a prática profissional.”.

Assim, a CI trabalha com as atividades e processos envolvidos no tratamento e na manipulação da informação, buscando otimizar o seu acesso por parte dos usuários.

A partir da concepção de que a CI se estabelece como uma ciência pós-moderna, explica-se a presença de múltiplas teorias que permeiam o campo (CAPURRO; HJØRLAND, 2003; NHACUONGUE; FERNEDA, 2015), bem como as divergências nas definições de seu objeto de estudo – a informação. Capurro e Hjørland (2003) estabelece a informação como algo físico, cognitivo ou social, de acordo com o contexto vivido. Similarmente, Silva e Gomes (2015) reconhecem a

informação como epistemológica, técnica (pragmática) ou social, enquanto Buckland (1991) introduz a informação como processo, conhecimento ou coisa.

Buckland (1991, p. 352) coloca que:

A intenção pode ser que os usuários se tornem informados (informação-como-processo) e que haja uma transmissão de conhecimento (informação-como-conhecimento). Mas os meios fornecidos, o que é manuseado e operado, o que é armazenado e recuperado, é informação física (informação-como-coisa).

Nesse sentido, a informação com a qual o profissional da informação é responsável de manipular é a informação como coisa, aquela quantificável, mensurável, sendo tudo aquilo que pode ser considerado como informativo para uma determinada comunidade, em determinado momento, como dados, textos e documentos, eventos e objetos (BUCKLAND, 1991). Segundo Buckland (1991), a informação como coisa é uma evidência: ela atesta alguma coisa, em algum período; trata-se de uma qualidade atribuída às coisas, que “[...] denota algo relacionado à compreensão, algo que, se encontrado e entendido corretamente, poderia mudar o conhecimento, as crenças, sobre alguma questão.” (BUCKLAND, 1991, p. 353, tradução nossa).

Segundo Smit (2012), a informação na CI deve ser registrada e institucionalizada. Dessa forma, a informação deve ser inscrita em algum suporte que seja passível de manipulação para que os profissionais da informação consigam manejá-la. Esse ponto é também corroborado por Nhacuongue e Ferneda (2015, p. 14), segundo os quais “[...] a Ciência da Informação trabalha com a informação registrada ou conhecimento objetivo.” Além disso, aborda-se a necessidade da informação ser institucionalizada, para que receba um “selo de qualidade” que a caracterize como uma informação válida e de utilidade para alguém (SMIT, 2012, p. 86).

Nesse contexto, Bush (1945) coloca a relevância de se armazenar as informações, pensando-se em preservar os conhecimentos em memória de determinada comunidade, bem como para sua recuperação e uso posterior, introduzindo, ainda, o estabelecimento de associações entre informações relacionadas como forma de expandir o conteúdo que é recuperado pelo usuário. É a partir da recuperação da informação, conforme coloca Saracevic (2009), que é possível ter acesso à informação registrada, sendo o momento em que os usuários

conseguem obter materiais relevantes a eles. Compreende-se como relevante aquilo que é útil para responder os problemas apresentados pelos indivíduos, também chamado de informação “momentaneamente importante”, como coloca Bush (1945, n.p., tradução nossa).

O processo de recuperação da informação envolve diferentes dimensões, sendo: os aspectos cognitivos, a partir do momento em que a definição de relevância é uma atividade que decorre do processo cognitivo dos usuários; o contexto, tanto o ambiente informacional que provê o suporte informacional, quanto o social, que permite que o usuário formule suas perguntas; e as tecnologias que apoiarão o processo.

Essas tecnologias auxiliam o usuário a identificar o que é relevante ou não de maneira automatizada. Santos e Vidotti (2009) abordam a necessidade de se observar essas tecnologias de informação e comunicação (TIC) de forma completa, considerando o contexto social que as permeiam, visto que as tecnologias são criadas pela sociedade pensando-se no desenvolvimento dela, não devendo ser dissociadas uma da outra. Conforme as autoras colocam, “[...] sem tais ações essencialmente humanas, nenhuma das facilidades instrumentais oferecidas pelas novas tecnologias é realmente transformadora.” (SANTOS; VIDOTTI, 2009, n.p.).

Assim, as tecnologias devem ser vistas sob um ângulo diferente, de um ponto de vista em que se prioriza a sociedade, o usuário, e suas necessidades, devendo ser adequadas para melhor servi-la. Como Santos e Vidotti (2009) definem, trata-se de adotar “Um olhar diferente! Não porque se olhava errado, mas porque se pode olhar diferente.” (SANTOS; VIDOTTI, 2009, n.p.).

A partir da perspectiva do desenvolvimento da pesquisa científica, Jim Gray introduz que, atualmente, vive-se o início do quarto paradigma da ciência (CORDEIRO *et al.*, 2013). De modo geral, a ciência apresenta quatro paradigmas: a ciência empírica, visando a observação e descrição de fenômenos naturais; a ciência teórica, momento em que se tem a aplicação de modelos e generalizações; a ciência computacional, com base em simulações dos fenômenos; e, por fim, a *eScience*, decorrente do crescente e massivo número de dados gerados (CORDEIRO *et al.*, 2013; HEY; TANSLEY; TOLLE, 2009).

Conforme Sayão e Sales (2019, p. 35-36) colocam, “O quarto paradigma científico, ou *eScience*, é descrito na literatura como a reconfiguração dos paradigmas anteriores – ciência experimental, ciência teórica e ciência baseada em

simulação – em torno de um mundo rico em dados.”. Para Oliveira *et al.* (2015, p. 223, tradução nossa), o quarto paradigma

Essencialmente traz ciência aberta, dados científicos abertos e gerenciamento, compartilhamento, agregação, curadoria, preservação e cooperação científica para o uso e reutilização de pesquisas científicas. Nós estamos na era de dados intensivos.

Nesse sentido, em conjunto com a visão de Buckland (1991) de informação como coisa, é possível encontrar a justificativa da necessidade do envolvimento da CI com o tratamento de dados, assim como os dados de pesquisa. Dentre as diferentes atividades em que os profissionais da informação se inserem, menciona-se o processo de representação da informação, essencial para a recuperação da informação. Desse modo, é necessário fazer uso dos metadados. A partir da aplicação dos metadados na *Web* é possível realizar a identificação dos recursos e, portanto, a sua recuperação.

Com a expansão do número de informações disponibilizadas na sociedade, em conjunto com a expansão da percepção do que é considerado como informacional, presencia-se uma ampla gama de tipos de objetos considerados como informativos, seja de acordo com a sua natureza (texto, imagem, imagem em movimento etc.), com a estrutura de armazenamento que dá suporte informacional (livro, *e-book*, entrada em blog etc.) ou com o ambiente em que se encontram (analógico ou digital). Nesse sentido, observa-se a importância de se aplicar os processos, métodos e instrumentos de tratamento da informação, inseridos no âmbito da Ciência da Informação, a uma variedade maior de recursos informacionais, como forma de permitir que eles de fato possam ser acessados e utilizados pelos indivíduos.

Ao abordar questões de como possibilitar que a informação possa ser recuperada, apresenta-se como imprescindível a abordagem ao processo de representação da informação. Assim, reconhece-se a relevância de se compreender como esse processo se constitui. A representação tem origem na esfera da cognição do indivíduo, a partir da internalização de objetos do mundo exterior e da compreensão dos conceitos pela mente. A formação de uma representação passa por um processo de abstração; então, a partir de contato externo com o objeto, os

sentidos internos são ativados e a mente passa a conhecer e entender o objeto singular em questão (LANDIM FILHO, 2010).

Desse modo, é possível afirmar que a representação é formada pelo conhecimento que se tem do objeto, da presença de um conceito internalizado no indivíduo, bem como da imagem transitória e preliminar formada no intelecto, que Landim Filho (2010), retomando Tomás de Aquino, refere-se como 'fantasma'. Esse fantasma é a imagem formada pelo intelecto sobre um determinado objeto quando se tem algum estímulo externo e pode ocorrer antes mesmo de um conceito ser formado.

A partir desses processos cognitivos forma-se a representação interna na mente do indivíduo, não apenas uma imagem, mas sim a relação entre ela, o conceito e os conhecimentos possuídos. Essa concepção do intelecto é abordada por Castro (2008) e Marcondes (2001) como algo natural e inerente ao indivíduo – uma forma de assimilar o mundo exterior.

Assim, é possível expressar a representação interna como uma representação externa, remetendo a um recurso informacional, de forma que os usuários possam obter acesso e realizar a recuperação do mesmo, podendo a representação ser usada e reutilizada (CASTRO, 2008).

No contexto da Ciência da Informação pode-se dizer que representar consiste em descrever determinado recurso informacional de acordo com as suas principais características, de maneira que ele possa ser identificado em meio a outros recursos e, ainda, que permita ao usuário a compreensão do que se trata determinado recurso sem a necessidade de acessá-lo. Dessa maneira, a representação se apresenta como uma semelhança ao objeto original sendo representado, referindo-se a uma forma reduzida do mesmo (CASTRO, 2008; MARCONDES, 2001). Marcondes (2001) acrescenta que uma representação se forma, também, por informações adicionais que não constam no objeto original, como número de classificação, de forma a melhor servir o usuário em sua busca informacional.

Além do mais, objetiva-se potencializar o tempo do usuário para que o mesmo não precise analisar integralmente cada um dos possíveis recursos informacionais que viriam a satisfazer sua necessidade informacional, podendo recorrer às representações que permitirão ao usuário verificar todas as principais informações, de determinado(s) acervo(s), relativas ao recurso com maior praticidade.

Sendo a representação externa fruto de um processo cognitivo que ocorre em cada indivíduo, diretrizes e padrões devem ser adotados de forma a uniformizar as representações e obter consistência na organização da informação e propiciar a sua recuperação.

Observando a realidade vivida atualmente, em que a sociedade se encontra permeada por tecnologias e pelo crescimento da *Web*, essa preocupação com a uniformização torna-se ainda mais pertinente, frente ao obstáculo de auxiliar o usuário a identificar e obter as informações que deseja. Nesse sentido, cabe ao cientista da informação intervir com a aplicação apropriada da representação, com a utilização de metadados, como possível solução para a presente desordem que marca a *Web*, sendo que, dentre os diferentes tipos de recursos informacionais passíveis de serem representados, encontram-se os dados de pesquisa.

A utilização dos metadados data desde os primórdios da organização da informação, sendo que com o desenvolvimento da *Internet* e de ambientes tecnológicos os metadados passam a estar mais perceptíveis (ALVES, 2010; GILLILAND, 2016; CASTRO; SIMIONATO; ZAFALON, 2016). Nas palavras de Santos, Simionato e Arakaki (2014, p. 148), “Os recursos informacionais possuem características que requisitam um tratamento descritivo específico para otimizar sua busca, acesso e recuperação.”. Assim, conforme mencionam Zeng e Qin (2016, p. 12, tradução nossa), “[...] metadados existem não apenas no universo de dados bibliográficos tradicionais, mas também em nosso dia a dia.”, estando presente, por exemplo, em embalagens de variados produtos, como alimentos ou produtos de limpeza.

Eles são compreendidos como os dados que identificam os recursos informacionais, de forma a padronizá-los; são dados sobre dados (GILLILAND, 2016; CASTRO; SIMIONATO; ZAFALON, 2016), definido por Pomerantz (2015, p. 26, tradução nossa) como “[...] uma declaração sobre um objeto potencialmente informativo.”. O termo ‘dado’, apesar de bastante abordado e discutido, não apresenta uma definição única e unânime na literatura (CÓRDULA; ARAÚJO, 2019). De modo geral, os dados podem ser compreendidos como unidades básicas que, quando contextualizadas, possuem potencial informativo.

Santos e Sant’Ana (2015, p. 205) definem o dado a partir de uma perspectiva de modelagem de dados, para os quais o

[...] dado é uma unidade de conteúdo necessariamente relacionada a determinado contexto e composta pela tríade entidade, atributo e valor, de tal forma que, mesmo que não esteja explícito o detalhamento sobre contexto do conteúdo, ele deverá estar disponível de modo implícito no utilizador, permitindo, portanto, sua plena interpretação.

Gilliland (2016, n.p., tradução nossa) coloca que os metadados podem ser vistos como “[...] a soma total do que se pode dizer em um determinado momento sobre qualquer objeto de informação em qualquer nível de agregação.”. Conforme Rautenberg *et al.* (2018, p. 32) estabelecem, “Eles contribuem para gerar informações de modo que os humanos e as máquinas possam encontrar, entender e usar os dados ou um conjunto de dados e suas distribuições.”. Ou seja, os metadados possuem como objetivo a descrição dos recursos informacionais, de modo que seja possível identificar e acessar os mesmos (GILLILAND, 2016; RAUTENBERG *et al.*, 2018).

Alves (2010, p.47, grifo do autor) define o conceito de metadados da seguinte maneira:

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.

Desse modo, além de proporcionar o acesso aos recursos (GILLILAND, 2016; RAUTENBERG *et al.*, 2018), a utilização de metadados permite a localização, a busca, e, conseqüentemente, a recuperação da informação (ALVES, 2005; 2010). Nesse sentido, Simionato (2012) ressalta que a essencialidade dos metadados se apoia na indispensabilidade da realização do registro do recurso, proporcionando a organização, a interatividade, a validação, a recuperação e a preservação da informação. Zeng e Qin (2016) apontam, ainda, a promoção da avaliação e do gerenciamento das entidades a partir do uso de metadados.

Ao tratar dos metadados abordam-se três componentes: o esquema, o elemento e o valor. O esquema é um conjunto de regras que dita quais tipos de declarações podem ser feitas, enquanto o elemento é o atributo, a unidade dos metadados que indica qual dado sobre o recurso deve ser apontado, como ‘título’, ‘criador’ etc. Já o valor “[...] é o dado que é designado a um elemento [...]”, extraído

de um recurso, como 'Leonardo da Vinci' em 'criador: Leonardo da Vinci' (POMERANTZ, 2015).

Os metadados são adotados para explicitar as características dos recursos informacionais, como “[...] estrutura, conteúdo, qualidade, contexto, origem, propriedade e condição.” (SANTOS; SIMIONATO; ARAKAKI, 2014, p. 150). Dessa forma, eles são categorizados em algumas tipologias, sendo que estas variam de acordo com os diferentes autores que os abordam. A exemplo, para Riley (2017) os metadados podem ser administrativos (subdivididos em técnicos, de preservação e de direitos), descritivos, estruturais ou linguagens de marcação, enquanto Gilliland (2016) categoriza os metadados como administrativos, descritivos, de preservação, técnicos e de uso.

Os metadados administrativos oferecem informações para a gestão dos recursos informacionais (GILLILAND, 2016; RILEY, 2017), ligados também à sua criação (RILEY, 2017). Arakaki (2019, p. 67) dispõe exemplos de metadados administrativos, sendo eles:

[...] data de criação dos recursos, tipos de arquivos, formas de acesso, controle de direitos e reproduções, informação sobre registros legais, informação sobre localização, o nome da instituição que cria os objetos digitais, a data da digitalização, o equipamento de digitalização usado, o nome do arquivo digital mestre, o arquivo de exibição e o arquivo de miniatura, dados do ciclo de vida da informação, como datas de criação de arquivos digitais, revisões subsequentes, tempo de revisão para retenção, arquivamento ou descarte, nomes criadores e revisores do recurso e níveis de autorização para uma determinada função, informação de aquisição e avaliação; direitos e acompanhamento de reprodução; documentação de requisitos e protocolos legais, culturais e de acesso à comunidade; informação de localização; critérios de seleção para digitalização; documentação de repatriamento digital, são exemplos de metadados administrativos, aquisição de informações, direitos e rastreamento de reprodução, requisitos legais de acesso e informações de localização.

Os metadados técnicos, que para Riley (2017) são um tipo de metadado administrativo, estão “[...] relacionados a como um sistema funciona ou metadados se comportam.” (GILLILAND, 2016, n.p., tradução nossa); são “[...] informações sobre arquivos digitais necessários para decodificar e renderizá-los, como tipo de arquivo [...]” (RILEY, 2017, n.p., tradução nossa). Já os metadados de preservação visam a preservação e a gestão a longo prazo dos recursos informacionais. Os

metadados de direitos, apontado como categoria por Riley (2017, n.p., tradução nossa), “[...] detalham os direitos de propriedade intelectual ligados ao conteúdo.”

Os metadados descritivos são “[...] utilizados para identificar, autenticar e descrever coleções e recursos de informação confiáveis relacionados.” (GILLILAND, 2016, n.p., tradução nossa), sendo necessários para compreender os recursos (RILEY, 2017).

Riley (2017, n.p., tradução nossa) aponta para a existência dos metadados estruturais, que “[...] descrevem os relacionamentos de partes de recursos entre si [...]”, como as páginas de um livro em sequência. Ainda segundo Riley (2017, n.p., tradução nossa), tem-se as linguagens de marcação, em que “Essas linguagens misturam metadados e conteúdo juntos, uma prática usada apenas algumas vezes com outras formas de metadados.”. Trata-se da integração entre metadados e conteúdo pensando-se na sinalização de características de determinado conteúdo.

Gilliland (2016, n.p., tradução nossa) ainda estabelece os metadados de uso, definidos pela autora como “Metadados relacionados ao nível e tipo de uso de coleções e recursos de informação.”, como “Registros de circulação; Registros de exposições físicas e digitais; Rastreamento de uso e usuário; Reuso de conteúdos e informações de multiversão; Registros de pesquisa; Metadados de direito.” (GILLILAND, 2016, n.p., tradução nossa).

O panorama das diferentes categorias atribuídas pelos diferentes autores na literatura pode ser melhor observado no quadro de Arakaki (2019) (Quadro 1):

Quadro 1 – Categorias de metadados na literatura

Autor	Baseado em	Categorias						
Gilliland (1999, 2008, 2016)	-	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Greenberg (2001)	-	Administrativo	Descoberta	-	-	Uso - Técnicos - Direitos	-	Autenticação
Méndez Rodríguez (2002)	-	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Senso e Rosa Piñero (2003)	Gilliland-Swetland (1999)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Riley (2004)	-	Administrativo - Técnico - Direitos - Preservação	Descritivo	-	-	-	Estrutural	-
Haynes (2004)	Gilliland (1998)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Hillmann ; Marker; Brady (2008)	-	Administrativo	Descritivo	Preservação	-	Acesso/Uso	Estrutural	-
Alves (2010)	Gilliland-Swetland (1999), Senso e Rosa Piñero (2003) e Rosetto (2003)	Administrativo	Descritivo	Conservação	Técnicos	Uso	-	-
Miller (2011)	-	Administrativo - Técnico e preservação - Direitos - Uso	Descritivo	-	-	-	Estrutural	-
Alves e Santos (2013)	Gilliland (2008), Gilliland-Swetland (1999), Senso e Rosa Piñero (2003) e Rosetto (2003)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Pomerantz (2015)	-	Administrativo - Metadados técnicos - Estruturais - Proveniência - Preservação - Direitos - Meta-metadata	Descritivo	-	-	Uso - Data exhaust - Paradata	-	-
Zeng e Qin (2008; 2016)	Gilliland (2008)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Gartner (2016)	-	Administrativo - Técnico - Direitos - Preservação	Descritivo	-	-	-	Estrutural	-
Riley (2017)	-	Administrativo - Técnico - Direitos - Preservação	Descritivo	-	-	-	Estrutural	Linguagem de marcação
Haynes (2017)	Gilliland (2016)	Administrativo	Descritivo	Preservação	Técnicos	Uso	-	-
Joudrey e Taylor (2018)	-	Administrativo - Técnico, - Preservação - Direitos e acesso - Meta-metadata	Descritivo	-	-	-	Estrutural	-

Fonte: Arakaki (2019).

A partir desse quadro observa-se que muitos dos autores se baseiam nos trabalhos de Gilliland (1998; 2008; 2016) e Gilliland-Swetland (1999). Além disso, é possível identificar que muitas das categorias de metadados estabelecidas por diferentes autores são convergentes, sendo que em diversos casos eles são condensados sob a categoria de metadados administrativos, como é o caso de Riley (2004), Miller (2011), Pomerantz (2015), Gartner (2016), Riley (2017) e Joudrey e Taylor (2018), levando à compreensão de que, para os autores, esses tipos de metadados compartilham características e aplicabilidades administrativas (ARAKAKI, 2019).

Dessa forma, os metadados apresentam papel central na organização de dados publicados, sendo os responsáveis pela padronização e recuperação desses dados. Para uniformizar a aplicação dos elementos de metadados podem ser utilizados os vocabulários, de modo a padronizar os termos empregados.

3 OS DADOS DE PESQUISA

É de amplo conhecimento que o desenvolvimento e a publicação de pesquisas científicas são imprescindíveis para o progresso das diferentes áreas do conhecimento, e conseqüentemente, do país que as suportam, motivo pelo qual vigoram inúmeros programas de incentivo e fomento à pesquisa. Nacionalmente, citam-se o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) e as Fundações de Amparo à Pesquisa (FAPs) (BRASIL, 2017).

Dessa maneira, discute-se sobre a importância do fornecimento de acesso aos dados que geram as pesquisas, uma vez que são instrumentos chave para o desenvolvimento científico, podendo servir de base para a realização de novas pesquisas (GRAY *et al.*, 2002). Esses dados, chamados de dados de pesquisa, consistem em fontes primárias que embasam as pesquisas e investigações científicas, sendo necessários para a validação de suas descobertas e de seus resultados (THE CASRAI DICTIONARY, 2015; ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, 2007; SAYÃO; SALES, 2015).

Para a *Organisation for Economic Co-Operation and Development* (2007), “Um conjunto de dados de pesquisa constitui uma representação sistemática e parcial do assunto investigado.” Os dados “[...] podem ser dados experimentais, dados observacionais, dados operacionais, dados de terceiros, dados do setor público, dados de monitoramento, dados processados ou dados adaptados.” (THE CASRAI DICTIONARY, 2015, n.p., tradução nossa). Semelhantemente, Semeler e Pinto (2019, p. 115) apontam que os “[...] dados de pesquisa científica são o resultado de qualquer investigação sistemática que envolva processos de observação, experimentação ou simulação de procedimentos de pesquisa científica.”. Assim sendo, existem diversos tipos de dados e formas de administrá-los, de acordo com cada pesquisa e campo científico ao qual pertencem (SILVA SEGUNDO; ARAÚJO, 2019).

De acordo com a *Research Information Network* (2008), os dados de pesquisa podem ser gerados ou coletados pensando-se no benefício dos integrantes de determinado projeto, ou de grupos de pesquisadores ou usuários mais amplos em uma certa disciplina, de forma a prover dados e informações de referência. Os dados podem ser produzidos em dois distintos estágios de uma pesquisa: “[...] dados brutos criados ou reunidos no curso de experimentos ou observações; ou

dados derivados, resultantes do processamento ou combinação de dados “brutos” ou outros dados.”, sendo que ambos podem ser compartilhados (RESEARCH INFORMATION NETWORK, 2008, p. 14, tradução nossa). Sayão e Sales (2015) também apontam para os dados canônicos ou referenciais, sendo dados já consolidados, como o caso de dados correspondentes à sequência genética ou às estruturas químicas.

Sayão e Sales (2015) também identificam e categorizam os tipos de dados de pesquisa de acordo com sua origem e natureza, alguns dos quais são identificados pelo *The Casrai Dictionary* (2015), porém sem essa distinção, conforme visto anteriormente. Quanto à origem, os dados podem ser observacionais, computacionais ou experimentais. Os dados observacionais consistem em dados que só podem ser diretamente observados uma única vez, sendo provenientes de eventos que ocorreram em determinado local, data e hora, como “[...] a erupção de determinado vulcão numa data específica, a fotografia de uma supernova, o levantamento das atitudes de uma comunidade [...]” (SAYÃO; SALES, 2015, p. 7-8). Os dados computacionais “São resultados da execução de modelos computacionais ou de simulações [...]” (SAYÃO; SALES, 2015, p. 8), enquanto os dados experimentais são aqueles relacionados a experimentos científicos executados em situações controladas.

Já de acordo com a natureza, os dados de pesquisa podem ser: “Números, imagens, vídeos ou áudio, software, algoritmos, equações, animações ou modelos e simulações.” (SAYÃO; SALES, 2015, p. 8).

A preservação e o armazenamento dos dados de pesquisa devem ser feitos para que eles possam auxiliar no desenvolvimento de futuras pesquisas, seja ao se basear em dados históricos, como os dados observacionais, ou para replicar situações e modelos anteriormente realizados a partir de dados computacionais e experimentais.

A *Research Information Network* (2008, p. 26, tradução nossa) menciona algumas motivações para a publicação de dados de pesquisa, sendo:

[...] o altruísmo e ações para o bem do aprendizado; [...] a cultura de compartilhamento de dados [...]; o aumento da visibilidade das instituições e grupos de pesquisa; as oportunidades para coautoria de artigos [...]; as oportunidades de colaboração com outros, dentro e fora do nicho de assunto [...]; o feedback positivo, estimado pelas

agências de fomento [...]; o encorajamento dos pares; e a própria especialização e interesse em questões relacionadas a dados.

A *Research Information Network* (2008) aborda alguns incentivos para estimular o compartilhamento dos dados de pesquisa pelos pesquisadores, sendo: a evidenciação dos benefícios que a publicação dos dados pode apresentar; mecanismos de citação padronizados e a explicitação das recompensas que o pesquisador obteria, como qual seria o reconhecimento de se publicar os dados de pesquisa ou se influenciaria na obtenção de subsídios.

Um exemplo de iniciativa em direção a melhor publicação de dados de pesquisa é o *DataCite*, cujo objetivo consiste em prover identificadores do tipo *Digital Object Identifier* (DOI) para a identificação única de conjuntos de dados de pesquisa (DATACITE, [2020?]).

Ao publicar os dados de uma pesquisa aumenta-se a visibilidade dos cientistas, promove-se a sustentabilidade a partir do reuso dos dados e melhora-se a utilização dos mesmos, conseqüentemente culminando no avanço da ciência (COSTELLO, 2009; GRAY *et al.*, 2002; SMITH, 2009). Maiores discussões acerca da utilização e reutilização de dados de pesquisa são feitas por Curty *et al.* (2017), Curty (2019), Dias, Anjos e Araújo (2019), dentre outros.

O compartilhamento desses dados consiste em uma boa prática, tendo sido cobrado pela sociedade e apoiado por políticas governamentais como forma de melhorar a “[...] responsabilidade e transparência da ciência.” (SMITH, 2009, p. 1, tradução nossa). Costello (2009) aponta, ainda, como motivo da demanda da publicação de dados o fato de que estes são diretamente ou indiretamente sustentados por financiamentos públicos ou, então, direcionados ao público.

No tópico do uso intensivo de dados, Laender *et al.* (2020, n.p.) dizem que

Espera-se que a capacidade de vinculação e interseção de conjuntos de dados provenientes de fontes diferentes aumente a precisão, o poder preditivo e a generabilidade das descobertas científicas e ajude os pesquisadores a identificar futuras direções de investigação. A disponibilidade de grandes volumes de dados fornece um incentivo à busca por novos procedimentos e ferramentas computacionais para obtenção, armazenamento, organização e análise desses dados, podendo trazer aperfeiçoamentos na forma de produzir e de dar maior transparência a todo o processo de criação do conhecimento científico.

Apesar dos benefícios e motivações apresentados, os pesquisadores encontram-se desencorajados a publicar seus dados diante alguns obstáculos, como a falta de recompensa ou reconhecimento acadêmico, o tempo gasto e trabalho envolvido e a falta de padronizações para guiar o processo (GRAY *et al.*, 2002). A *Research Information Network* (2008, p. 26-28, tradução nossa) apresenta uma lista de limitações indicadas pelos pesquisadores:

- falta de tempo e de recursos;
- falta de tempo para lidar com requisições de acesso aos dados e às informações;
- falta de experiência ou conhecimento em gestão de dados;
- falta de conhecimento e dificuldade em tornar os dados disponíveis, acessíveis e utilizáveis;
- restrições legais ou éticas, principalmente em casos de trabalhos colaborativos e pesquisas que envolvem dados pessoais, como em casos de entrevistas;
- falta de conhecimento quanto a como arquivar os dados, tanto para escolher o melhor repositório ou banco de dados quanto para publicar por si próprios;
- fatores competitivos que levam os pesquisadores a quererem ter acesso exclusivo aos seus dados ou então ter controle sobre quem pode ter acesso aos mesmos;
- medo de exploração ou uso indevido dos dados, como a interpretação errada dos dados, de forma diferente da que o autor pretendia, ou o medo de que alguém vá se apropriar dos dados após realizar uma mudança mínima aos mesmos;
- descrença de que alguém vá achar os dados úteis;
- recompensas limitadas ou não especificadas.

Assim, torna-se necessário o envolvimento e incentivo das instituições e agências de fomento, que devem “[...] facilitar e encorajar a publicação e reuso de dados de pesquisa [...]” (RESEARCH INFORMATION NETWORK, 2008, p. 10, tradução nossa), mostrando os benefícios do processo, por meio de trabalhos que tratam sobre o tema, provendo apoio na publicação de dados, e criando um sistema

de recompensas acadêmicas para aqueles que publicam os dados de suas pesquisas (RESEARCH INFORMATION NETWORK, 2008). Nota-se que o desenvolvimento e uso de padrões e modelos podem auxiliar a manter a consistência dos conjuntos de dados, ajudando na questão de uniformização (COSTELLO, 2009).

Kim e Stanton (2013) identificaram fatores que influenciam o compartilhamento de dados pelos pesquisadores, envolvendo motivações pessoais e institucionais, corroborando com os pontos já abordados, sendo identificado que: a pressão reguladora por parte dos periódicos, a pressão normativa em nível disciplinar, o benefício para a carreira e o altruísmo individual se relacionam de modo positivo com o comportamento de compartilhamento de dados de pesquisa (KIM; STANTON, 2013). Já o fator “esforço percebido”, ou seja, pesquisadores que percebem o esforço requerido para compartilhar dados, possui relação negativa (KIM; STANTON, 2013). Entretanto, os fatores de pressão reguladora por agências de fomento, a disponibilidade de repositórios de dados e o risco de carreira percebido não apresentaram relações significativas com o compartilhamento de dados (KIM; STANTON, 2013).

Quanto às características atuais do compartilhamento de dados de pesquisa, Santos e Sant’Ana (2019, p. 65) resumem a oposição que existe quando o assunto é o compartilhamento desses dados, dizendo que:

Os interesses públicos tendem a justificar a abertura do acesso aos dados o quanto antes como parte da contrapartida de recursos investidos na pesquisa e em prol do bem comum. Os princípios éticos, por outro lado, reforçam a necessidade de preocupações como as voltadas para garantias de direitos como privacidade e de transparência na realização de pesquisas para garantir que os limites morais e éticos não sejam desrespeitados.

Ressalta-se que diferentes áreas do conhecimento possuem perspectivas distintas sobre os dados referentes às pesquisas, seja sobre o tipo de dados que geram e coletam ou quanto à cultura de compartilhamento dos dados de pesquisa (CARVALHO; LEITE, 2019).

Nesse contexto, com a *Web* a publicação dos dados de pesquisa torna-se cada vez mais viável, sendo seu compartilhamento *online* uma forma de expor as produções do pesquisador para um público ainda maior (COSTELLO, 2009).

Algumas áreas de pesquisa já adotam a publicação de dados como atividade fundamental, como a Astronomia, a Física e a Climatologia, com a emergência, ainda, de revistas direcionadas à publicação de dados (COSTELLO, 2009; GRAY *et al.*, 2002; RESEARCH INFORMATION NETWORK, 2008).

Destaca-se que, nacionalmente, algumas iniciativas têm sido desenvolvidas na esfera de compartilhamento de dados de pesquisa, como a parceria entre o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) e a Rede Nacional de Ensino e Pesquisa (RNP) para o desenvolvimento de uma rede brasileira de dados de pesquisa (LEANDER *et al.*, 2020).

Menciona-se também a Rede de Repositórios de Dados Científicos do Estado de São Paulo³, um metabuscador de dados de pesquisa lançado em dezembro de 2019 pela FAPESP, em parceria com: a Universidade Estadual de Campinas (Unicamp), a Universidade de São Paulo (USP), a Universidade Estadual Paulista (Unesp), a Universidade Federal de São Carlos (UFSCar), a Universidade Federal do ABC (UFABC), a Universidade Federal de São Paulo (Unifesp), o Instituto Tecnológico de Aeronáutica (ITA) e a Embrapa Informática Agropecuária (CNPTIA/Embrapa) (LEANDER *et al.*, 2020; ZIEGLER, 2019). A plataforma, desenvolvida pela USP, permite a busca de dados de pesquisa mantidos pelas instituições a partir de um único buscador, ampliando o alcance e visibilidade dos dados (LANDER *et al.*, 2020; ZIEGLER, 2019).

Ademais, acrescenta-se que a publicação de dados de pesquisa e a sua gestão são abordadas, ainda, por autores como Curty e Aventurier (2017), Moreno (2018), Ribeiro (2019), dentre outros.

Smit, Van der Hoeven e Giaretta (2011), ao abordarem a preservação digital e o envolvimento de editoras, apontam para a importância da descrição dos dados de pesquisa pelo emprego de metadados para garantir a preservação deles, conforme resumem em uma das estratégias de preservação digital:

Metadação semântica: mantendo as descrições do objeto digital e garantindo que eles possam ser entendidos pela comunidade de usuários alvo no futuro - isso é frequentemente o mecanismo mais importante para dados científicos, pois inclui o conhecimento tácito sobre o objeto. (SMIT; VAN DER HOEVEN; GIARETTA, 2011, p. 40, tradução nossa, grifo do autor).

³ Disponível em: <https://metabuscador.uspdigital.usp.br/>.

Indo além, Smit, Van der Hoeven e Giaretta (2011, p. 41, tradução nossa) dizem que “A metadação adequada tornou-se um elemento mais importante mais recentemente, como um passo vital para a compreensão e autenticidade futuras.”. Isso mostra o papel central que os metadados, assim como o controle deles, possuem na publicação dos dados de pesquisa e em sua preservação.

Segundo Sayão e Sales (2015, p. 11), “Os dados e as coleções de dados de pesquisa possuem um tempo de vida maior que os projetos de pesquisa que os criaram.”, sendo necessário considerar o tratamento que deve ser dado aos mesmos para que seja garantido o seu uso. Para tanto, um modelo de gestão do ciclo de vida dos dados deve ser adotado para guiar o manuseio dos dados de pesquisa, como o *Digital Curation Centre (DCC) Curation Lifecycle Model*, o *DataONE Data Lifecycle*, o *DDI Combined Lifecycle Model*, o *UK Data Archive Data Lifecycle*, dentre outros (SAYÃO; SALES, 2015). Como coloca Córdula e Araújo (2019, p. 197), “A função essencial do gerenciamento de dados é de apoiar e facilitar o compartilhamento e o reuso dos dados.”, sendo que “O ciclo de vida dos dados é um processo contínuo de desenvolvimento, manipulação, gerenciamento e estágios de armazenamento de dados.” (CÓRDULA; ARAÚJO, 2019, p. 198).

Nas palavras de Silva Segundo e Araújo (2019, p.131, grifo do autor),

[...] considerando que as tecnologias da informação e da comunicação possibilitam ampla produção, acesso e distribuição de dados digitais, gerir dados, objetivando acesso livre, é mais do que simplesmente adotar qualquer plano de gestão e instalar um *software* de processamento e de análise. O gestor, aqui tratado como curador, tomará decisões baseadas no ciclo de vida dos dados digitais, objetivando utilizá-lo adequadamente ao longo dos anos, para prevenir perdas e contaminações.

A curadoria opera tanto a curto prazo, ao favorecer a manutenção dos dados, quanto a longo prazo, pensando-se em sua preservação (RESENDE; BAX, 2019; SILVA SEGUNDO; ARAÚJO, 2019). Ainda no contexto das tecnologias de informação e a gestão de dados, Sant’Ana (2019, p. 14) afirma que elas “[...] têm se transformado em fator-chave nos ciclos de vida dos dados, indo além de suportar a coleta e armazenamento, e participar efetivamente do tratamento, do filtro, da seleção e, até mesmo, geração de novos conteúdos [...]”.

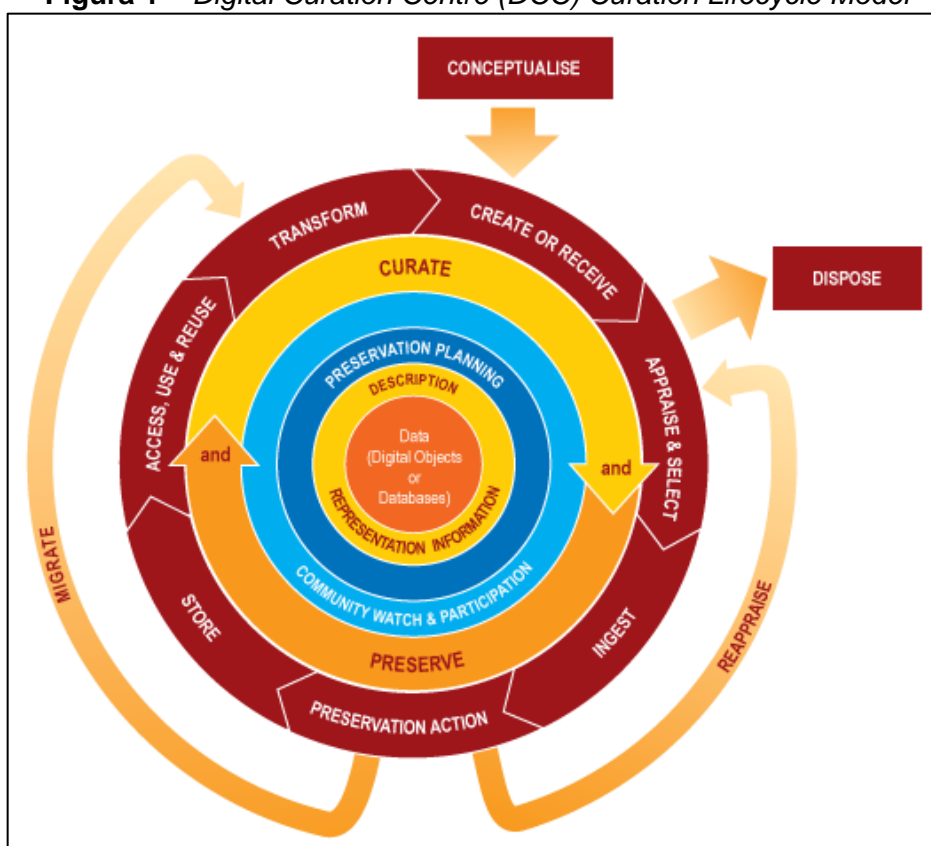
Para melhor compreender o que um modelo de ciclo de vida de dados envolve, optou-se por apresentar o *DCC Curation Lifecycle Model*, devido ao seu

alto nível de detalhamento (DIGITAL CURATION CENTRE, c2019c). Em seguida, apresenta-se brevemente o ciclo de vida do *Research360*, voltado ao contexto de pesquisas, com foco institucional, (SILVA SEGUNDO; ARAÚJO, 2019), sendo abordado por trazer em maiores detalhes as atividades e os agentes envolvidos.

O DCC *Curation Lifecycle Model* foi desenvolvido pelo *Digital Curation Centre* (DCC), criado em 2004 com o objetivo de solucionar problemas na esfera da curadoria digital (DIGITAL CURATION CENTRE, c2019d), é “[...] um centro de especialização reconhecido internacionalmente em curadoria digital, com foco na criação de capacidades e habilidades para gestão de dados de pesquisa.” (DIGITAL CURATION CENTRE, c2019a, n.p., tradução nossa). Entende-se por curadoria digital todo o processo de gestão e as atividades que garantem qualidade aos recursos digitais, envolvendo “[...] manter, preservar e agregar valor aos dados de pesquisa digitais durante todo o seu ciclo de vida.” (DIGITAL CURATION CENTRE, c2019e, n.p., tradução nossa).

O DCC *Curation Lifecycle Model* é um modelo que “[...] fornece uma visão geral gráfica de alto nível dos estágios necessários para a curadoria bem-sucedida e a preservação de dados a partir da conceituação inicial ou do recebimento através do ciclo de curadoria iterativa.” (DIGITAL CURATION CENTRE, c2019c, n.p., tradução nossa), conforme a Figura 1.

Figura 1 – Digital Curation Centre (DCC) Curation Lifecycle Model



Fonte: Digital Curation Centre (c2019c).

O modelo é composto pelas ações de ciclo de vida todo (*full lifecycle actions*), as ações sequenciais (*sequential actions*) e as ações ocasionais (*occasional actions*). As ações de ciclo de vida todo, correspondente à parte central do modelo, são ações presentes durante todo o ciclo de vida do dado, sendo:

- **descrição e representação da informação** (*description and representation information*): atribuição de metadados, com o uso de padrões, necessários para compreender os dados e mantê-los a longo prazo;
- **planejamento de preservação** (*preservation planning*): estabelecimento dos planos de gestão das ações de curadoria do ciclo de vida;
- **vigilância e participação na comunidade** (*community watch and participation*): observação das atividades da comunidade e participação no desenvolvimento de padrões, ferramentas e *software*;
- **curar e preservar** (*curate and preserve*): cuidado e administração das ações planejadas para a curadoria e preservação dos dados durante seu ciclo de vida (DIGITAL CURATION CENTRE, c2019c, n.p., tradução e grifo nosso).

As ações sequenciais, presentes no modelo em vermelho, são as que ocorrem de modo cíclico, sendo:

- **conceituar** (*conceptualise*): concepção e planejamento da criação dos dados, incluindo métodos de captura e opções de armazenamento;
- **criar ou receber** (*create or receive*): criação dos dados, incluindo metadados administrativos, descritivos, estruturais, técnicos e de preservação, ou recepção dos dados de criadores de dados, outros arquivos, repositórios ou centros de dados, incluindo a atribuição de metadados caso necessário;
- **avaliar e selecionar** (*appraise and select*): avaliação dos dados e seleção para curadoria e preservação a longo prazo;
- **ingerir** (*ingest*): transferência dos dados para guarda, como um arquivo, repositório ou centro de dados;
- **ação de preservação** (*preservation action*): realização de ações de preservação a longo prazo de modo a garantir a autenticidade, confiabilidade, usabilidade e integridade dos dados (limpeza, validação, atribuição de metadados de preservação, atribuição de informações de representação e garantia de estruturas de dados ou formatos de arquivo aceitáveis);
- **armazenar** (*store*): armazenamento dos dados de forma segura, aderindo a padrões;
- **acesso, uso e reuso** (*access, use and reuse*): garantir que os dados sejam sempre acessíveis para usuários e reutilizadores designados;
- **transformar** (*transform*): criação de novos dados a partir dos dados originais, como por meio de migração a um formato diferente ou pela criação de um subconjunto dos dados para criar novos resultados derivados (DIGITAL CURATION CENTRE, c2019c, n.p., tradução e grifo nosso).

As ações ocasionais são as que ocorrem de maneira eventual, de acordo com a situação:

- **descarte** (*dispose*): descarte dos dados que não farão parte do processo de curadoria, podendo ser transferido para outro armazenamento ou destruído;

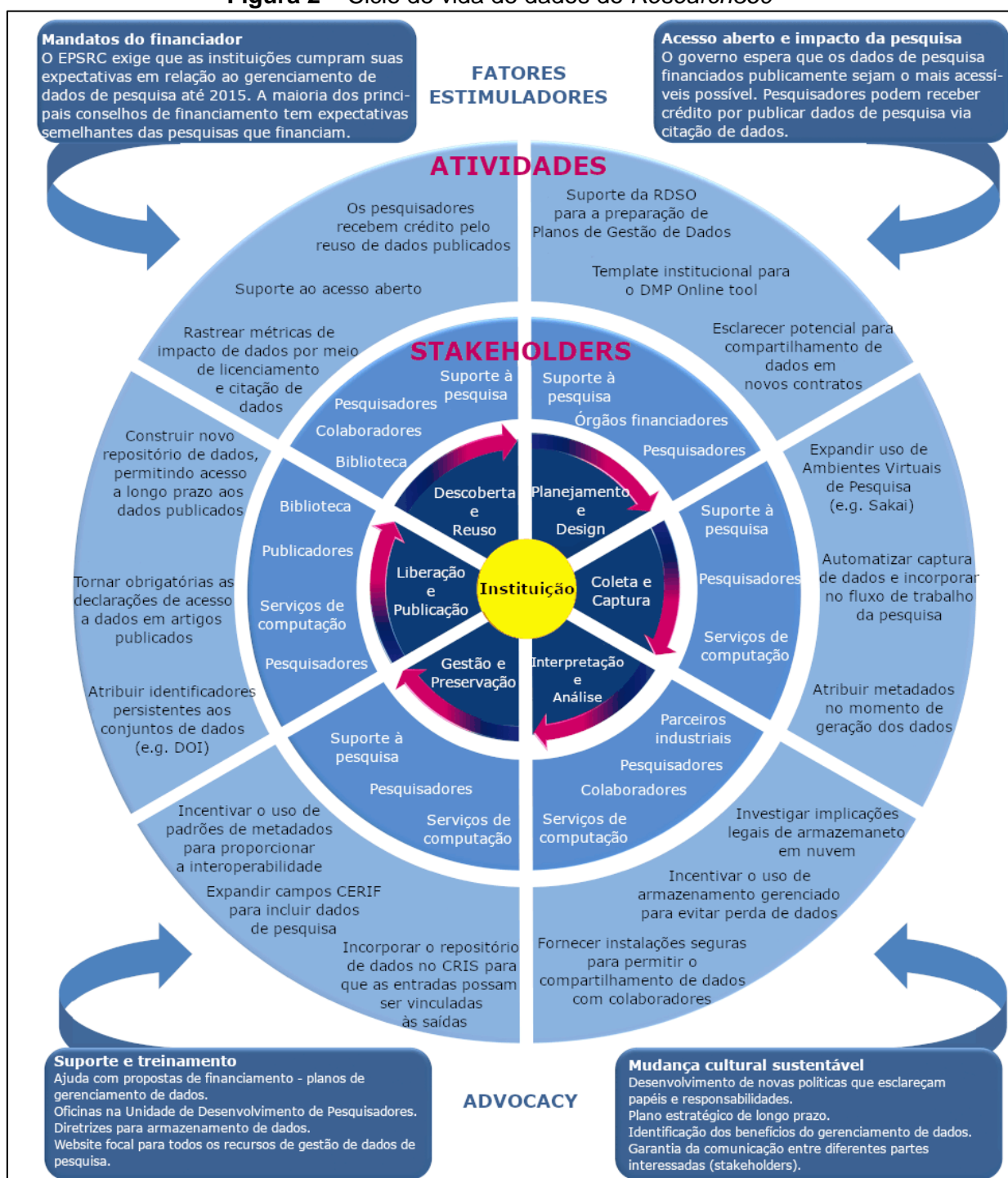
- **reavaliação** (*reappraise*): retorno dos dados que falharam no processo de validação para uma nova avaliação e reseleção;
- **migração** (*migrate*): migração dos dados para um formato diferente, ocorrendo em decorrência de novo acordo para o armazenamento ou para proteger os dados de obsolescência de *hardware* e *software* (DIGITAL CURATION CENTRE, c2019c, n.p., tradução e grifo nosso).

Outro ciclo de vida a se abordar é o pertencente ao projeto *Research360*, realizado de outubro de 2011 a julho de 2013, a partir de uma parceria entre o Centro de Formação Doutoral em Tecnologias Químicas Sustentáveis (*Doctoral Training Centre in Sustainable Chemical Technologies*), da Universidade de Bath (*University of Bath*), o DCC e a *Neil Beagrie of Charles Beagrie Ltd.* (UKOLN, 2013).

Foi um grande projeto institucional destinado a pilotar políticas, infraestrutura e recursos de treinamento para fornecer suporte de alta qualidade à pesquisa, para garantir que a Universidade de Bath e seus parceiros obtenham o melhor valor possível dos dados gerados pelas pesquisas. (UKOLN, 2013, n.p., tradução nossa).

De acordo com Silva Segundo e Araújo (2019), o modelo apresentado pelo projeto tem como “[...] principal objetivo [...] envolver a instituição durante o gerenciamento dos dados.”, abordando os papéis da instituição em cada uma das etapas. Assim, o ciclo possui, em seu centro, a instituição, conforme pode ser visualizado na Figura 2 a seguir.

Figura 2 – Ciclo de vida de dados do Research360



Fonte: tradução de McKen *et al.* (2012).

O ciclo do *Research360* apresenta em seu núcleo a instituição, seguida das etapas que ocorrem ciclicamente, das partes interessadas (*stakeholders*) e das atividades compreendidas em cada uma de suas etapas. Além disso, apresenta os fatores que estimulam (*drivers*) as atividades, bem como a defesa (*advocacy*) – ou o

porquê – da gestão dos dados de pesquisa (MCKEN *et al.*, 2012), conforme pode ser visto na Figura 2, que apresenta a tradução do ciclo de McKen *et al.* (2012).

Tendo em vista que o ciclo foi apresentado em 2012, no contexto de um projeto desenvolvido por instituições europeias, o modelo apresenta algumas características específicas, como: 1) o atendimento às expectativas de gerenciamento de dados do *Engineering and Physical Sciences Research Council* (EPSRC) até 2015; e 2) o apoio da *Research Designs & Standards Organisation* (RDSO) na fase de planejamento e *design*. Todavia, com exceção da data imposta, as demais particularidades podem ser compreendidas em um contexto maior, entendendo que 1) existe uma cobrança e expectativas por parte das agências financiadoras no geral, e que 2) o suporte à pesquisa deve se envolver na preparação de planos de gestão de dados (PGDs).

Segundo Sayão e Sales (2016), o cenário científico encontra-se em mudança, com o desenvolvimento de novos instrumentos e métodos, apresentando mudanças nas relações entre os pesquisadores e nas formas de compartilhamento e comunicação científicos. Nesse contexto,

os dados deixam de ser um mero subproduto da atividade de pesquisa e se tornam protagonistas na geração de novos conhecimentos e descobertas. Para tal, precisam estar disponíveis e interpretáveis para que possam transmitir conhecimento no tempo e no espaço e para que sejam reusados em diversos contextos alimentando a pesquisa interdisciplinar. (SAYÃO; SALES, 2016, p. 68).

Assim, os modelos de ciclo de vida dos dados permitem que os dados, nesse caso os dados de pesquisa, sejam de qualidade e assim permaneçam a longo prazo, de forma que os usuários possam acessar e utilizar dados autênticos, confiáveis e íntegros. Outros autores, como Torino, Vodotti e Sant’Ana (2019) e Araújo *et al.* (2019), abordam outros ciclos de vida de dados voltados ao contexto dos dados de pesquisa, buscando compreender melhor os processos e atores envolvidos, bem como o envolvimento da área de Ciência da Informação.

No contexto da Ciência da Informação temos que a responsabilidade sobre o processo de curadoria dos dados de pesquisa pode ser atribuído ao profissional da informação. De acordo com Semeler e Pinto (2019, p. 123), o tratamento desses dados figura na chamada “Biblioteconomia de dados”, que

[...] corresponde ao interesse de bibliotecários em compreender atividades ligadas ao gerenciamento e à curadoria de todos os tipos de dados, sendo seu foco o tratamento, a gestão e a curadoria de dados de pesquisas em qualquer disciplina científica.

A função do bibliotecário de dados é também abordada por Tartarotti, Dal'Evedove e Fujita (2019, p. 452), para as quais

[...] os bibliotecários de dados tornam-se especialistas em informação científica, capazes de reunir e assessorar no desenvolvimento de coleções de conjuntos de dados de pesquisa, uso e preservação dos dados, como também no acesso a dados já arquivados ou na criação de novos dados, como parte de sua prática de pesquisa.

Compreende-se, assim, que o bibliotecário é o profissional apto a trabalhar com os dados, encontrando-se envolvido com os procedimentos de gestão e curadoria de dados de pesquisa, bem como com as atividades ligadas ao seu desenvolvimento e em sua utilização, a partir da aplicação dos conhecimentos e práticas tradicionais da Biblioteconomia e da Ciência da Informação (SEMELER; PINTO, 2019; TARTAROTTI; DAL'EVEDOVE; FUJITA, 2019). Semeler e Pinto (2019, p. 116) ressaltam também que os dados de pesquisa “[...] são parte da ciberinfraestrutura que deve ser investigada para geração de inovação em bibliotecas.”, de modo que a atuação do profissional da informação esteja condizente com o cenário informacional, que se encontra sempre em desenvolvimento.

Pensando na utilização dos conhecimentos biblioteconômicos, a *Research Data Alliance* (RDA) lista 23 recursos gratuitos para auxiliar na atividade de gestão de dados de pesquisa, introduzindo-os nas práticas do bibliotecário (RESEARCH DATA ALLIANCE, 2016). Esses 23 elementos encontram-se dispostos em nove diferentes categorias, sendo elas: recursos de aprendizagem, referência e divulgação de dados, planos de gerenciamento de dados, literacia de dados, citação de dados, licenciamento e privacidade de dados, preservação digital, repositórios de dados, além de uma comunidade de prática (RESEARCH DATA ALLIANCE, 2016).

Em cenário europeu, Tenopir *et al.* (2016) buscaram identificar os serviços de dados de pesquisa que são oferecidos pelas bibliotecas europeias membros da *Ligue des Bibliothèques Européennes de Recherche* (LIBER) a partir de questionários (*survey*). Os autores verificaram que a maioria das bibliotecas oferecem ou planejam oferecer serviços de consulta, como “[...] serviços como a

criação de guias web, serviços de referência e/ou consultoria com professores e alunos [...]” (TENOPIR *et al.*, 2016, p. 2, tradução nossa), limitando-se no que diz respeito a serviços técnicos de dados de pesquisa, “[...] tal como criar e manter um repositório de dados e/ou preparar conjuntos de dados para inclusão em um repositório de dados.” (TENOPIR *et al.*, 2016, p. 2, tradução nossa). Os possíveis motivos para essa configuração são o tempo e recursos demandados da instituição e de seus membros, além de maiores habilidades técnicas para tal (TENOPIR *et al.*, 2016). Assim, demonstra-se a necessidade de que os profissionais da informação e as instituições se engajem com o cenário dos dados de pesquisa, visando o aprimoramento e atualização de suas capacidades e competências.

O planejamento e gestão dos dados de pesquisa é uma etapa necessária para que seja possível acompanhar o seu envolvimento nas pesquisas. Van den Eyden *et al.* (2011, p. 6, tradução nossa), no guia de boas práticas para a gestão e o compartilhamento de dados do UK *Data Archive*, estabelecem que

Um plano de gestão e compartilhamento de dados ajuda os pesquisadores a considerar, quando a pesquisa está sendo projetada e planejada, como os dados serão geridos durante o processo de pesquisa e compartilhado depois com a comunidade de pesquisa mais ampla.

Sayão e Sales (2015, p. 15) detalham o PGD como um documento formal que descreve quais “[...] dados serão processados, coletados ou gerados; quais as metodologias e padrões que serão utilizados nesses processos; se, como e sob que condições esses dados serão compartilhados e/ou tornados abertos para a comunidade de pesquisa; e como eles serão curados e preservados.” Strasser *et al.* (2012) ressaltam que a devida descrição dos dados é essencial para a gestão de dados, de modo que os pesquisadores possam compreender e utilizar os dados.

De modo similar, o *Digital Curation Centre* (c2019b, n.p., tradução nossa) introduz que “Os planos geralmente indicam quais dados serão criados e como, e descrevem os planos de compartilhamento e preservação, observando o que é apropriado, dada a natureza dos dados e quaisquer restrições que possam precisar ser aplicadas.”. Basicamente, um PGD busca garantir a preservação e utilidade futura dos dados de pesquisa (CÓRDULA; ARAÚJO, 2019).

Observa-se que um PGD deve ser delineado no momento preliminar da pesquisa, adaptando-o durante o seu desenvolvimento, caso necessário (SAYÃO;

SALES, 2015; VAN DEN EYDEN *et al.*, 2011). Conforme Van den Eyden *et al.* (2011, p. 6, tradução nossa) colocam, “Implementar medidas de gestão de dados durante os estágios de planejamento e desenvolvimento da pesquisa evitarão pânico e frustração.”, uma vez que se torna mais fácil realizar a pesquisa e garantir a usabilidade dos dados provenientes dela com um plano do que deverá ser realizado com eles.

Em nível nacional, a FAPESP requer que os dados de pesquisa oriundos de projetos financiados por ela sejam compartilhados, visando auxiliar na promoção do avanço da ciência (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2014; [2018?]). A Fundação cobra, desde 2017, o PGD no momento de submissão de projetos de pesquisa temáticos⁴ (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, 2017).

A FAPESP especifica que o PGD deve ser um documento textual de até duas páginas que deve responder a duas questões: “1. Quais dados serão gerados pelo projeto, e 2. Como serão preservados e disponibilizados, considerando questões éticas, legais, de confidencialidade e outras.” (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, [2018?], n.p.). Deve conter, ainda, a descrição dos dados e metadados oriundos do projeto, as restrições legais ou éticas envolvidas na disponibilização dos dados, uma política que determine as condições de preservação e compartilhamento dos dados e a descrição dos mecanismos, formatos e padrões que devem ser utilizados para fornecer os dados (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, [2018?], n.p.).

Apesar das instruções para a construção do PGD poderem variar de acordo com a área do conhecimento, o conteúdo que deve constar no documento é em geral o mesmo (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, [2018?]; STRASSER *et al.*, 2012). A Fundação de Amparo à Pesquisa do Estado de São Paulo ([2018?]) lista alguns manuais que podem ser seguidos para elaboração de um plano, como os da *National Science Foundation* (EUA), do *National Institutes of Health* (EUA), do *National Endowment for the Humanities* (EUA),

⁴ “A modalidade **Auxílio à Pesquisa - Projeto Temático** destina-se a apoiar propostas de pesquisa com objetivos suficientemente ousados, que justifiquem a duração de até cinco anos e as condições especialmente favorecidas de apoio e financiamento, incluindo a possibilidade de concessão de solicitações complementares vinculadas ao projeto.” (FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO, [2019?], grifo do autor).

do *Data Curation Centre*, do programa *Horizon 2020*, além de manuais de universidades e artigos.

Sayão e Sales (2015), com base em elementos recomendados pelo *DataONE*, *Jisc*, *DCC* e *Inter-university Consortium for Political and Social Research (ICPSR)*, estabelecem um roteiro para ajudar na construção de um PGD, consistindo em determinar:

- os tipos de dados que serão produzidos;
- a quantidade de dados que serão gerados;
- como os dados serão coletados;
- como os dados serão processados;
- quais formatos serão utilizados;
- como os arquivos serão nomeados;
- quais serão as medidas de de garantia de controle de qualidade;
- se há coleções de dados disponíveis que servem para a pesquisa;
- se serão utilizados dados já existentes;
- como os dados serão mantidos a curto prazo e
- quem será o responsável pela gestão dos dados a curto prazo.

Ademais, existem ferramentas que auxiliam no desenvolvimento de um PGD, como o *DMPTool*⁵, do Centro de Curadoria da Universidade de Califórnia (*University of California Curation Center*), e o *DMPonline*⁶, do *DCC*, ambos contando com modelos de planos de diferentes instituições e diversas áreas do conhecimento, sendo de acesso livre. Citam-se como outros exemplos o *DMP Editor*⁷, do *National and Regional Climate Adaptation Science Centers*, o *IEDA DMP Tool*⁸, da *Interdisciplinary Earth Data Alliance (IEDA)*, e o *ezDMP*⁹, baseado no *IEDA DMP Tool*. O *IEDA DMP Tool* e o *ezDMP* também são de acesso livre, enquanto o *DMP Editor* possui acesso restrito.

⁵ Disponível em: <https://dmptool.org/>.

⁶ Disponível em: <https://dmponline.dcc.ac.uk/>.

⁷ Disponível em: <https://my.usgs.gov/dmpeditor/>.

⁸ Disponível em: <https://www.iedadata.org/dmp/>.

⁹ Disponível em: <https://ezdmp.org/>.

Além da FAPESP, outras agências e instituições instituem guias e políticas destinadas à gestão e ao compartilhamento de dados de pesquisa. Monteiro e Lucas (2019) investigaram as instituições financiadoras de pesquisa nacionais a fim de identificar a presença de políticas de compartilhamento de dados de pesquisa, em específico, o compartilhamento aberto. Dentre as 15 agências consideradas por elas, apenas cinco apresentaram políticas para compartilhamento de dados de pesquisa de modo aberto. As particularidades de cada agência e suas políticas constam no Quadro 2 a seguir.

Quadro 2 – Cobertura de políticas para compartilhamento de dados de pesquisa de instituições financiadoras nacionais

Financiador	Caráter das Políticas de Dados Abertos	Cobertura da política		Estipulações da Política					Suporte Fornecido			
		Resultados Publicados	Dados	Prazos	Plano de Dados	Acesso/ Compartilhamento	Curadoria de longo prazo	Monitoramento	Orientação	Repositório	Centro de Dados	Custos
CNPq												
FAPESP												
FAPEMIG												
Embrapa												
FIOCRUZ												

Legenda: Cobertura Total Cobertura Parcial Nenhuma Cobertura
 Recomendatória Mandatória Recomendatória/Mandatória

Fonte: Monteiro e Lucas (2019).

O Quadro 2 acima permite visualizar que as principais agências que apresentam políticas de compartilhamento de dados de pesquisa são o CNPq, a FAPESP, a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) e a Fundação Oswaldo Cruz (FIOCRUZ). A pesquisa de Monteiro e Lucas (2019) demonstra que as agências de fomento ainda se mostram tímidas quanto às suas políticas de dados de pesquisa.

4 PRINCÍPIOS *LINKED DATA* E O COMPARTILHAMENTO DE DADOS

A discussão dos princípios *Linked Data* remonta à *Web Semântica*, também chamada de *Web dados*, (HEATH; BIZER, 2011; WORLD WIDE WEB CONSORTIUM, c2015b), que consiste em uma proposta de extensão da *Web* de documentos, buscando proporcionar um ambiente formado por dados e informações estruturadas de forma que as máquinas possam processá-las automaticamente (BERNERS-LEE; HENDLER; LASSILA, 2001).

Nas palavras de Peset, Ferrer-Sapena e Subirats-Coll (2011, p. 165, tradução nossa), “A diferença entre a web de hipertexto e a web semântica é que enquanto uma vincula páginas ou documentos em HTML, a segunda vincula dados estruturados.”.

O *World Wide Web Consortium* (c2015b, n.p., tradução nossa), em consonância com Berners-Lee, Hendler e Lassila (2001), coloca que “O objetivo final da *Web de dados* é permitir que os computadores façam um trabalho mais útil e desenvolvam sistemas que possam suportar interações confiáveis na rede.”.

Como Santarém Segundo (2014, p. 3865) estabelece,

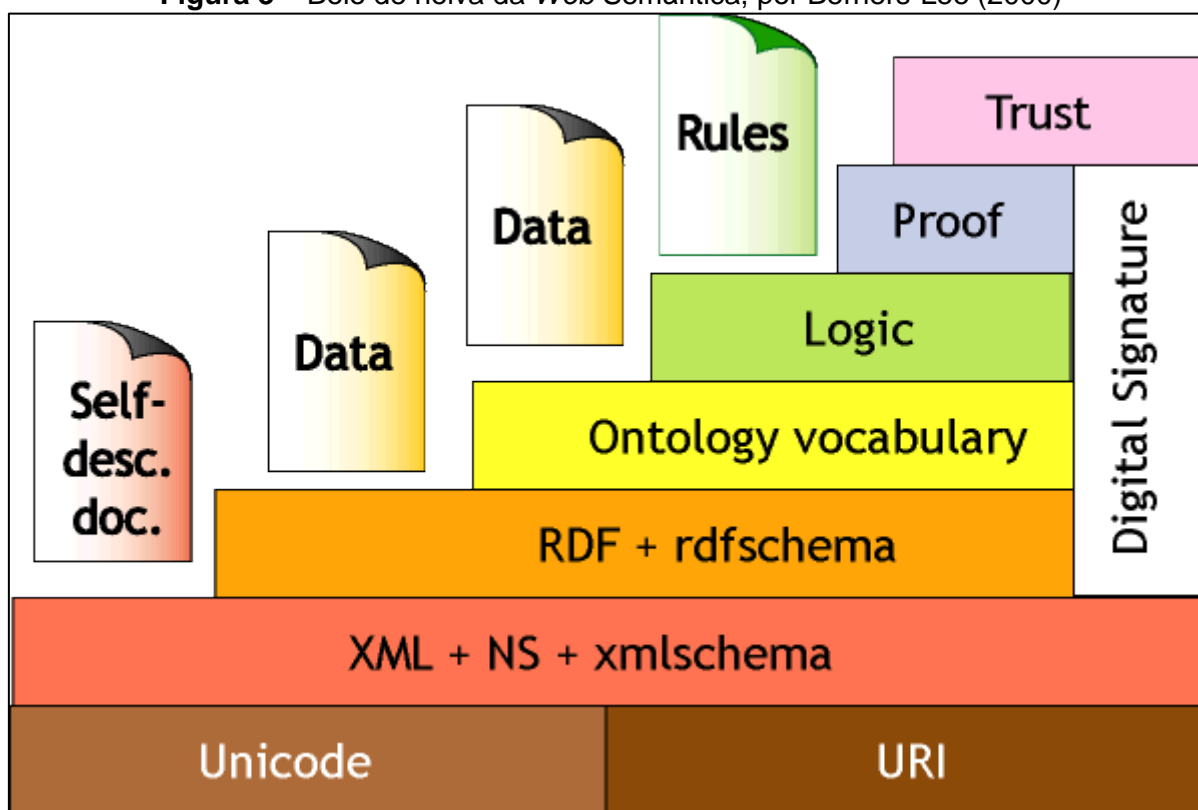
O projeto da *Web Semântica* tem como ponto fundamental a criação de uma nova estrutura de armazenamento de dados. O ponto principal está na separação da apresentação do conteúdo e do conteúdo da estrutura, tratando as unidades atômicas de uma informação como componentes independentes.

Para que as máquinas consigam assimilar e distinguir semanticamente os diferentes dados, algumas tecnologias fazem-se necessárias no processo. Essas tecnologias e outros elementos compõem a arquitetura da *Web Semântica*, por vezes apresentada graficamente como “bolo de noiva” (*layer cake*) (ARAKAKI, 2016; ISOTANI; BITTENCOURT, 2015).

A estrutura fora apresentada, inicialmente, por Berners-Lee (2000). A partir de então, diferentes versões foram criadas com base na estrutura do autor a fim de acompanhar o desenvolvimento da *Web Semântica*. Dentre as versões, citam-se as apresentadas por Nowack (2009) e Idehen (2017).

Na Figura 3 encontra-se presente a estrutura proposta por Berners-Lee (2000):

Figura 3 – Bolo de noiva da *Web Semântica*, por Berners-Lee (2000)

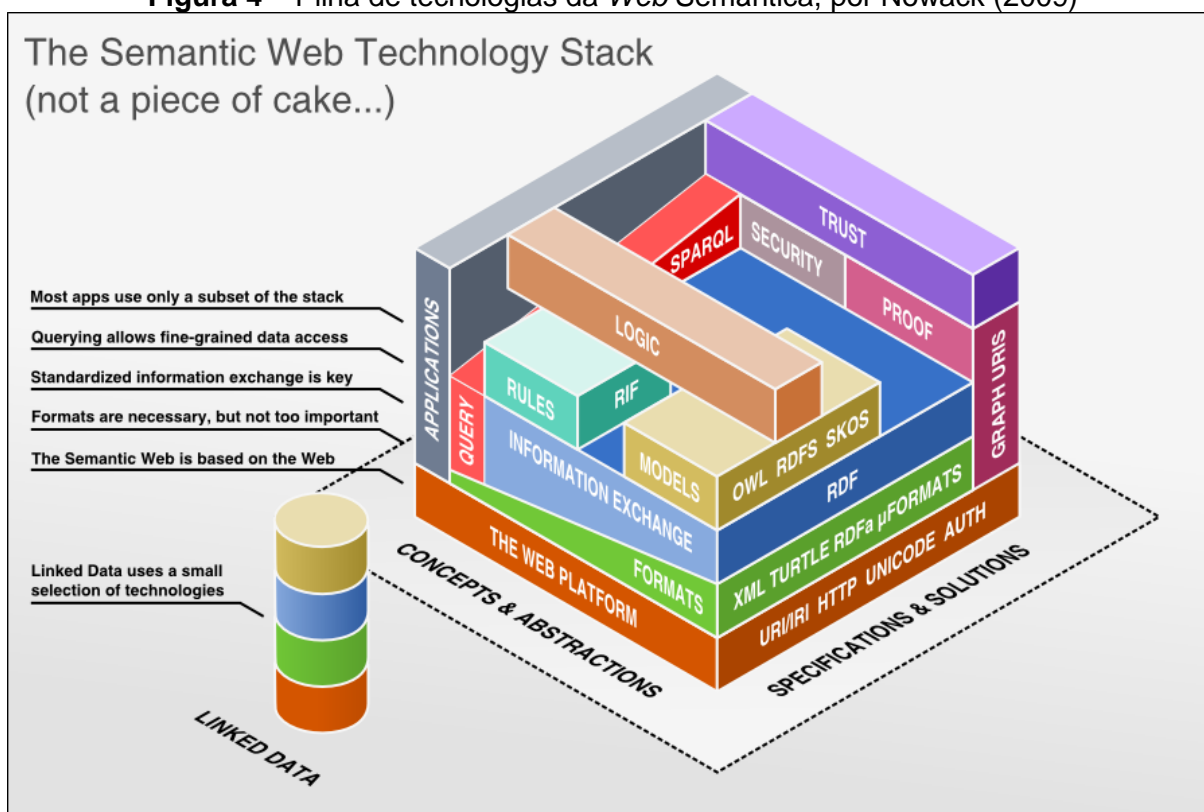


Fonte: Berners-Lee (2000).

Em sua arquitetura, Berners-Lee (2000) apresenta as diferentes tecnologias em camadas, em ordem crescente de poder (HORROCKS; PATEL-SCHNEIDER, 2003), com: UNICODE e URI na base, seguidos de XML, NS e XML Schema; RDF e RDF Schema; ontologias; lógica; prova e confiança.

A versão de Nowack (2009) apresenta a arquitetura como uma “pilha” (*stack*), conforme a Figura 4 a seguir:

Figura 4 – Pilha de tecnologias da *Web Semântica*, por Nowack (2009)

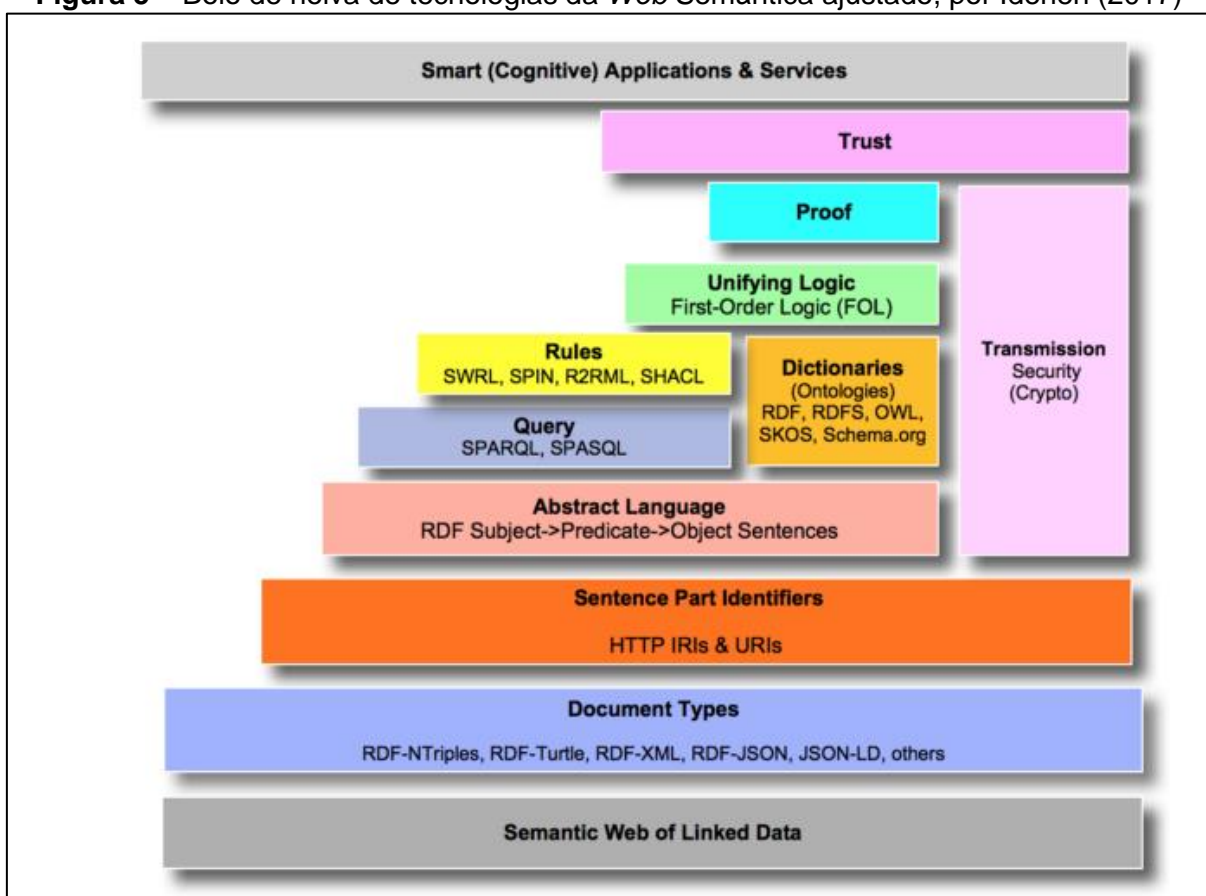


Fonte: Nowack (2009).

Essa estrutura coloca de forma mais detalhada e complexa cada uma das tecnologias, situando, ainda, os princípios *Linked Data*. Em sua constituição encontra-se a própria plataforma *Web* (URI/IRI, HTTP, UNICODE e AUTH), uma vez que a *Web Semântica* se baseia nela. Os formatos que a compõem são XML, Turtle, RDFa e uFORMATS. Para o intercâmbio de informação, Nowack (2009) indica o RDF, sendo que para a consulta aos dados indica o SPARQL. A utilização de modelos (OWL, RDFS e SKOS) também faz parte da arquitetura, como as regras (RIF), lógica, prova (*graph URIs*), segurança e confiança. O autor indica que a maior parte das aplicações utiliza apenas um subconjunto dos elementos dessa pilha.

Abaixo, na Figura 5, pode ser observada a versão de Idehen (2017):

Figura 5 – Bolo de noiva de tecnologias da *Web Semântica* ajustado, por Idehen (2017)



Fonte: Idehen (2017).

O bolo de noiva de Idehen (2017) incorpora explicitamente, assim como no modelo de Nowack (2009), as tecnologias que apoiam os princípios *Linked Data*, buscando ampliar a noção de formatos dos documentos. Desse modo, em sua base tem-se o que o autor chama de *Web Semântica do Linked Data*; em seguida, os tipos de documentos (RDF-NTriples, RDF-Turtle, RDF-XML, RDF-JSON e outros); os identificadores de partes de sentenças (HTTP IRIs e URIs); linguagem abstrata (sentenças RDF de sujeito>predicato>objeto); consulta (SPARQL, SPASQL); regras (SWRL, SPIN, R2RML, SHACL); dicionários, ou ontologias, (RDF, RDF *Schema*, OWL, SKOS, Schema.org); lógica unificadora (lógica de primeira ordem – FOL); prova; confiança; segurança de transmissão (criptografia); e aplicações e serviços inteligentes (cognitivos) (IDEHEN, 2017).

Nesse contexto, conforme observado, os princípios *Linked Data* fazem uso das tecnologias semânticas para se concretizarem, auxiliando na promoção de ligação de dados.

Traduzido ao português como dados ligados, dados interligados, dados vinculados ou dados conectados, o *Linked Data* fora proposto por Berners-Lee (2006) para otimizar o potencial informativo dos dados na *Web* ao melhorar a encontrabilidade de informações relacionadas, a partir da ligação de dados de forma compreensível por humanos e máquinas. Além disso, a ligação de diferentes comunidades promove a geração de novos conhecimentos (COYLE, 2016).

O *Linked Data* visa auxiliar na relação de dados presentes na *Web*, ajudando na semantificação do ambiente, uma vez que “A web semântica requer mais do que colocar dados estruturados online; ela requer a criação de ligações entre dados estruturados.” (POMERANTZ, 2015, p. 157, tradução nossa). O *World Wide Web Consortium* (c2015a, n.p., tradução nossa, grifo do autor) também defende a necessidade do *Linked Data* para que a *Web Semântica* se concretize, ao definir que “[...] não só a *Web Semântica* precisa de acesso aos dados, mas também devem estar disponíveis *relações entre os dados* para criar uma *Web de Dados* (em oposição a uma simples coleção de conjuntos de dados).”.

Conforme Heath ([201-?]) coloca, “*Linked Data* é sobre o uso da *Web* para conectar dados relacionados que não estavam vinculados anteriormente ou o uso da *Web* para reduzir as barreiras à ligação de dados atualmente ligados usando outros métodos.”.

De maneira similar, Santarém Segundo (2014, p. 3864) define *Linked Data* como “[...] um grande ambiente de informação através de dados ligados semanticamente entre si em vários lugares ao redor do mundo, caracterizando-se por utilizar a internet (a nuvem) como seu hospedeiro.”, podendo ser visto como os instrumentos e as ferramentas semânticas colocados em prática, perspectiva também apoiada por Coyle (2016).

De acordo com Arakaki (2016, p. 119), o *Linked Data* “[...] está [...] relacionado às melhores práticas para estruturação dos dados a partir das ferramentas e a garantia de troca de informações [...]”.

Nesse sentido, Berners-Lee (2006, n.p., tradução nossa) define os quatro princípios que estabelecem o comportamento esperado para que um dado possa ser considerado como *Linked Data*:

1. Uso de URIs como nome para as coisas;
2. Uso de HTTP URIs para que as pessoas possam procurar esses nomes;

3. Quando alguém procura um URI prover informações úteis utilizando padrões (RDF, SPARQL);
4. Incluir ligações a outros URIs, para que possam descobrir mais coisas.

Dessa forma, esse conjunto de princípios pode ser compreendido como recomendações ou boas práticas (ARAKAKI, 2016; BERNERS-LEE, 2006; PESET; FERRER-SAPENA; SUBIRATS-COLL, 2011) que guiam o que deve ser feito para que os dados na *Web* sejam ligáveis: 1) para um recurso pertencer à *Web Semântica* o mesmo deve ter um URI; 2) utilizar identificações do tipo HTTP URI, uma vez que “Eles fornecem uma maneira simples de criar nomes globalmente únicos de maneira descentralizada, pois todo proprietário de um nome de domínio ou delegado do proprietário do nome de domínio pode criar novas referências de URI.” e “Eles servem não apenas como um nome, mas também como um meio de acessar informações que descrevem a entidade identificada.” (HEATH; BIZER, 2011, tradução nossa, n.p.); 3) disponibilizar os dados em *Linked Data* de forma acessível, utilizando-se do RDF e do SPARQL, apresentando os dados em formatos como o RDF/XML; e 4) ligar os dados a outros dados na *Web* (BERNERS-LEE, 2006).

Quanto às tecnologias, os URIs são identificadores globais únicos atribuídos aos recursos para distingui-los dentre os demais, que podem ser do tipo *Uniform Resource Locator* (URL), que provê dados de localização do recurso, *Uniform Resource Name* (URN), que diz respeito ao nome atribuído a um determinado recurso, ou classificados como ambos (ISOTANI; BITTENCOURT, 2015). O *Hypertext Transfer Protocol* (HTTP) é o protocolo usado na “[...] transferência de documentos hipertextos [...]” (ISOTANI; BITTENCOURT, 2015, p. 58), enquanto a XML é uma linguagem de marcação recomendada pelo *World Wide Web Consortium* (W3C) para a representação e compartilhamento de informações estruturadas, utilizando-se de *tags* (etiquetas) customizáveis (WORLD WIDE WEB CONSORTIUM, 2015b). A XML permite que os recursos informacionais, bem como os dados e metadados que o descrevem, sejam melhor estruturados (SHIRADO; RAMALHO, 2019). Já o SPARQL é um protocolo e linguagem de consulta para RDF, “[...] para acesso aos dados.” (ISOTANI; BITTENCOURT, 2015, p. 33). De acordo com o Harris e Seaborne (2013, n.p., tradução nossa), “O SPARQL pode ser usado para expressar consultas em diversas fontes de dados, sejam eles armazenados nativamente como RDF ou visualizados como RDF a partir de um middleware.”.

A utilização do RDF, conforme colocam Heath e Bizer (2011, n.p., tradução nossa), “[...] visa ser empregado como uma língua franca, capaz de realizar a moderação entre outros modelos de dados usados na Web.”. Trata-se de um modelo que permite a publicação e o intercâmbio de dados na *Web* (FERREIRA; SANTOS, 2013; SCHREIBER; RAIMOND, 2014), de forma a descrever, mais precisamente, os recursos e as relações entre os recursos (ISOTANI; BITTENCOURT, 2015). Desse modo, o RDF é utilizado para realizar declarações sobre recursos, que são representadas por meio de triplas ‘sujeito-predicado-objeto’, sendo que o sujeito e o objeto são os recursos, enquanto o predicado indica o tipo de relação entre esses recursos (SCHREIBER; RAIMOND, 2014).

Existem três tipos de dados RDF que podem aparecer em uma tripla: IRIs, literais e nós vazios. Os *International Resource Identifiers* (IRIs) são uma generalização dos URIs, sendo identificadores que abrangem uma quantidade maior de caracteres do que os URIs (DUERST; SUIGNARD, 2005), podendo aparecer em um grafo como sujeito, predicado ou objeto em uma tripla (SCHREIBER; RAIMOND, 2014). Já os literais ocorrem na posição de objeto, sendo linhas de caracteres (*strings*), como nomes e datas, enquanto os nós vazios são utilizados para indicar recursos, sujeito ou objeto, que não possuam um identificador para expressar seus valores (SCHREIBER; RAIMOND, 2014).

Heath e Bizer (2011) identificam três principais tipos de ligações RDF, sendo ligações de relacionamento (*relationship links*), ligações de identidade (*identity links*) e ligações de vocabulário (*vocabulary links*).

As ligações de relacionamento ligam recursos que possuem alguma ligação entre si, como um autor a uma obra de sua autoria. Esses recursos possuem outras ligações, que podem levar a outros recursos relacionados e assim sucessivamente (HEATH; BIZER, 2011).

As ligações de identidade (*identity links*) apontam para os URIs que identificam um mesmo recurso, permitindo seu acesso e fornecendo maiores informações sobre ele. A importância dessas ligações cai sobre o fato de que os recursos possuem mais de um identificador, ou seja, mais de um URI, definido por diferentes entidades. Com o RDF pode-se estabelecer a afirmação de que determinados URIs correspondem a um mesmo recurso. Essas ligações são relevantes por três razões: (1) é possível observar as diferentes opiniões dos diferentes publicadores de dados; (2) permite associar os publicadores aos dados,

promovendo a rastreabilidade; (3) descomplica a identificação dos dados, pois a existência de um único identificador centralizaria a designação de URIs, e “A complexidade da coordenação, as despesas administrativas e burocráticas que isso introduziria criariam uma grande barreira ao crescimento na Web de dados.” (HEATH; BIZER, 2011, n.p., tradução nossa).

Já as ligações de vocabulário (*vocabulary links*) são essenciais para a integração dos dados, ligando os dados aos termos dos vocabulários, além de ligar os termos equivalentes de diferentes vocabulários. Trata-se de uma forma de lidar com a heterogeneidade promovida pela utilização de diferentes vocabulários para a descrição dos dados (HEATH; BIZER, 2011).

Com a adoção do RDF e dos demais recursos mencionados, torna-se possível publicar dados como *Linked Data*, proporcionando uma otimização na recuperação da informação, permitindo, ainda, a descoberta de novas fontes de informação e dados relacionadas à busca inicial, enriquecendo-a (BIZER; HEATH; BERNERS-LEE, 2009).

Bizer, Heath e Berners-Lee (2009, n.p., tradução nossa) mencionam os três passos a serem seguidos no processo de publicação de conjuntos de dados como *Linked Data*:

1. Designar URIs às entidades descritas pelo conjunto de dados e fornecer para a desreferenciação desses URIs o protocolo HTTP nas representações RDF.
2. Definir ligações RDF a outras fontes de dados na Web, para que os clientes possam navegar na Web de Dados como um todo seguindo as ligações RDF.
3. Fornecer metadados sobre dados publicados, para que os clientes possam avaliar a qualidade dos dados publicados e escolher entre diferentes meios de acesso.

De acordo com Bizer (2009), o que ele chama de ‘*Web de Linked Data*’, ou seja, uma *Web* com dados publicados como *Linked Data*, pode ser considerada uma camada que se encontra entrelaçada com a *Web* clássica, apresentando algumas características equivalentes:

- Qualquer pessoa pode publicar dados na Web de Linked Data.
- Links conectam entidades, criando um grafo de dados globais que abrange fontes de dados e permite a descoberta de novas fontes de dados.

- Os dados são auto-descritivos. Se uma aplicação encontra dados representados usando um vocabulário desconhecido, o aplicativo pode resolver os URIs que identificam termos de vocabulário para encontrar suas definições.
- A Web de Linked Data é aberta, o que significa que os aplicativos podem descobrir novas fontes de dados em tempo de execução seguindo links. (BIZER, 2009, p. 88, tradução nossa).

Essa visão da proximidade entre a *Web* clássica e do *Linked Data* também se encontra presente em Heath e Bizer (2011, n.p., tradução nossa), para os quais “A ideia básica do Linked Data é aplicar a arquitetura geral da World Wide Web à tarefa de compartilhar dados estruturados em escala global.”

Entende-se que a disponibilização de dados na *Web* como *Linked Data* se encontra próximo à realidade da *Web* de documentos atualmente predominante, sendo um aprimoramento da forma como a informação é tratada nessa a partir da adição do estabelecimento de relações.

A relevância da publicação de dados como *Linked Data* reside na criação de relações entre conjuntos de dados e informações a partir de contribuições de diferentes entidades e indivíduos, uma vez que “[...] permite que qualquer pessoa contribua com conhecimentos exclusivos em um formato que possa ser reutilizado e recombinação com a experiência de outras pessoas.” (BAKER *et al.*, 2011, n.p., tradução nossa), fator que tornam ricas essas conexões. Com a publicação de dados ligados aumenta-se o valor desses dados, bem como facilita-se “[...] a descoberta, o acesso e a integração [...]” (HEATH; BIZER, 2011, n.p., tradução nossa) dos mesmos.

O *Linked Data* aparece, ainda, como *Linked Open Data* (LOD), relacionado ao movimento de dados abertos (*Open Data*). Segundo *The Open Definition* ([2019?], n.p., tradução nossa), um dado ou conteúdo “Aberto significa que qualquer pessoa pode acessar, usar, modificar e compartilhar livremente para qualquer finalidade (sujeito, no máximo, a requisitos que preservem a procedência e a abertura)”.

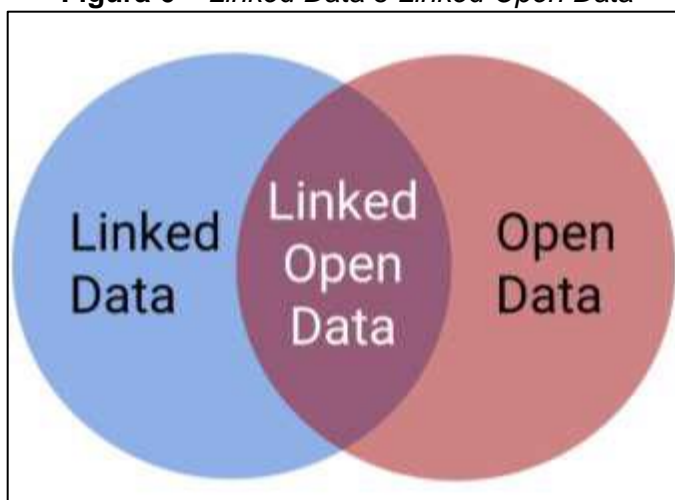
Compartilhar dados de maneira aberta é uma forma de promover a interoperabilidade entre diferentes ambientes e informações (OPEN KNOWLEDGE FOUNDATION, [2019?]). Conforme o *Open Knowledge Foundation* ([2019?], n.p., tradução nossa) afirma, “A interoperabilidade é importante porque permite que diferentes componentes trabalhem juntos. Essa capacidade de componentizar e ‘unir’ componentes é essencial para a construção de sistemas grandes e complexos.”

Laender *et al.* (2020, n.p.) ressaltam o benefício da disponibilização de dados abertos ao afirmarem que “A gestão e o compartilhamento desses dados vêm se tornando, cada vez mais, condição basilar para o aumento da cooperação internacional em pesquisa e, portanto, do progresso científico.”.

O *Linked Open Data* permite que se estenda o acesso e uso dos dados ligados, uma vez que se encontra livre de restrições de utilização. Segundo Peset, Ferrer-Sapena e Subirats-Coll (2011, p. 166, tradução nossa), “A web baseada em LOD é uma inovação na organização de conteúdo, que usa fontes de dados externas para criar novos serviços.”.

De forma sucinta, Berners-Lee (2006, n.p., tradução nossa) define que “Linked Open Data (LOD) é Linked Data liberado sob uma licença aberta, que não impede sua reutilização livremente.”. Ou seja, o *Linked Open Data* pode ser visto como a intersecção entre *Linked Data* e *Open Data* (Figura 6).

Figura 6 – Linked Data e Linked Open Data



Fonte: Arakaki (2016).

Berners-Lee (2006) define um sistema de estrelas para indicar em que ‘nível’ um dado em LOD se encontra, sendo que quanto maior o número de estrelas, mais fácil é para as pessoas utilizarem o(s) dado(s). Os níveis encontram-se descritos no Quadro 3.

Quadro 3 – Sistema de classificação por estrelas para *Linked Open Data*

★	Disponível na Web (qualquer formato), <i>mas com uma licença aberta, para ser Open Data</i>
★★	Disponível como dados estruturados legíveis por máquina (e.g. excel em vez de imagem digitalizada de uma tabela)

★★★	Como em (2), mais formato não proprietário (e.g. CSV em vez de excel)
★★★★	Todos os itens acima. Use padrões abertos do W3C (RDF e SPARQL) para identificar as coisas, para que as pessoas possam apontar para as suas coisas
★★★★★	Todos os itens acima, mais: ligue seus dados aos dados de outras pessoas para fornecer contexto

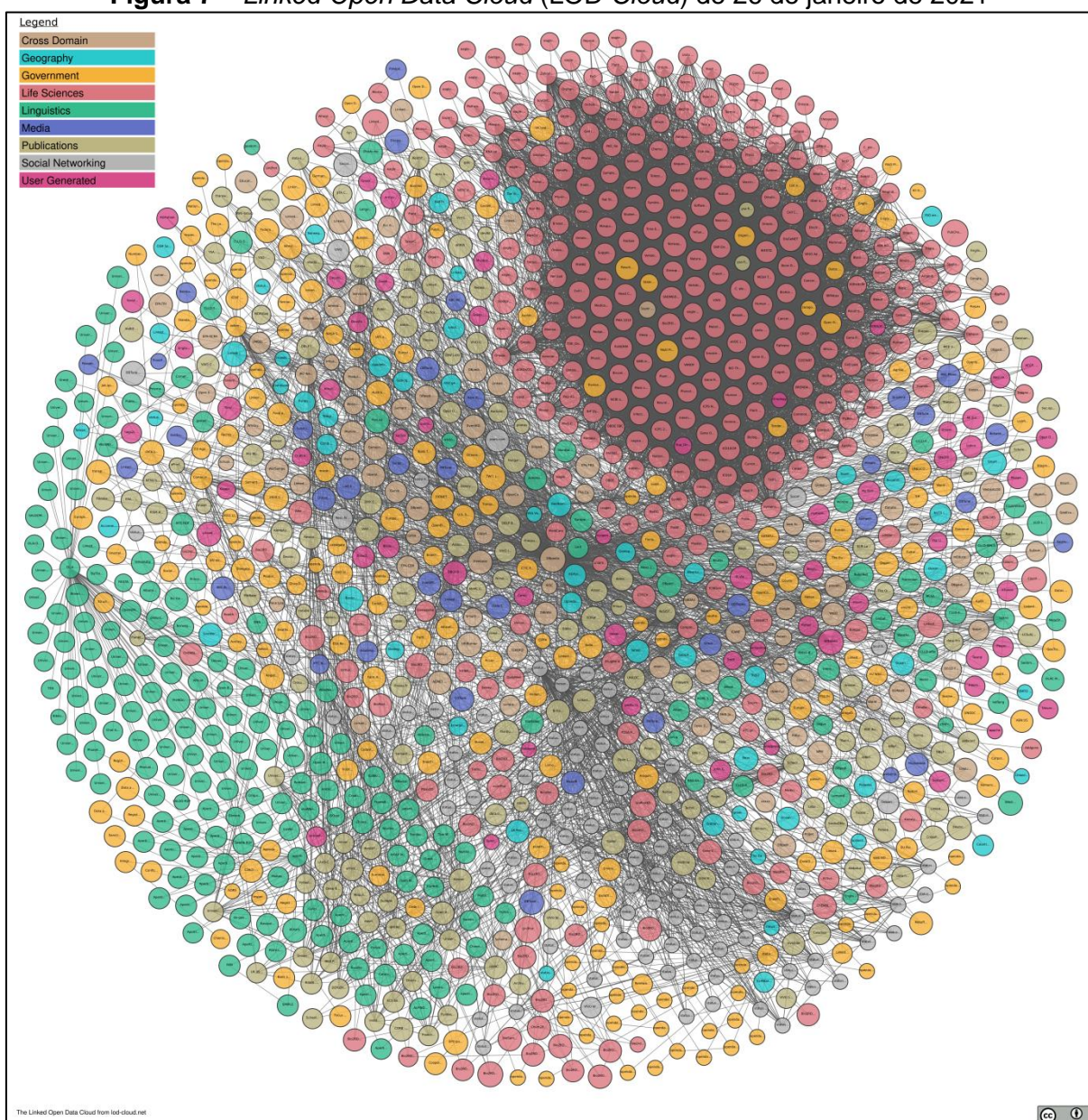
Fonte: tradução de Berners-Lee (2006, grifo do autor).

Assim, um dado pode ser considerado *Linked Data* caso esteja ligado a outros dados, respeitando os princípios que guiam sua publicação. Caso o dado como *Linked Data* estiver disponibilizado de forma a respeitar os critérios de *Open Data*, obtém-se dados como LOD.

Um dos esforços de aplicação do *Linked Data* é o projeto *Linking Open Data*, originado em 2007, tendo suporte do Grupo de Educação e Divulgação da Web Semântica do W3C (*W3C Semantic Web Education and Outreach Group – W3C SWEO*), cujo objetivo principal consiste em “[...] dar início à Web de dados, identificando os conjuntos de dados existentes que estão disponíveis sob licenças abertas, convertendo-os em RDF de acordo com os princípios *Linked Data* e publicando-os na Web.” (BIZER; HEATH; BERNERS-LEE, 2009, n.p., tradução nossa).

Os conjuntos de dados (*datasets*) participantes do projeto e as relações entre eles passaram a ser apresentados em forma de diagrama, chamado de *The Linking Open Data Cloud*, em maio de 2007, quando contavam com a participação de doze conjuntos de dados (BIZER; HEATH; BERNERS-LEE, 2009). Atualmente, com a nomenclatura alterada para *The Linked Open Data Cloud (LOD Cloud)*, o diagrama é mantido por John P. McCrae, do *Insight Centre for Data Analytics*, e encontra-se com um total de 1255 *datasets*, em sua versão de 26 de janeiro de 2021 (Figura 7) (MCCRAE *et al.*, [2021?]).

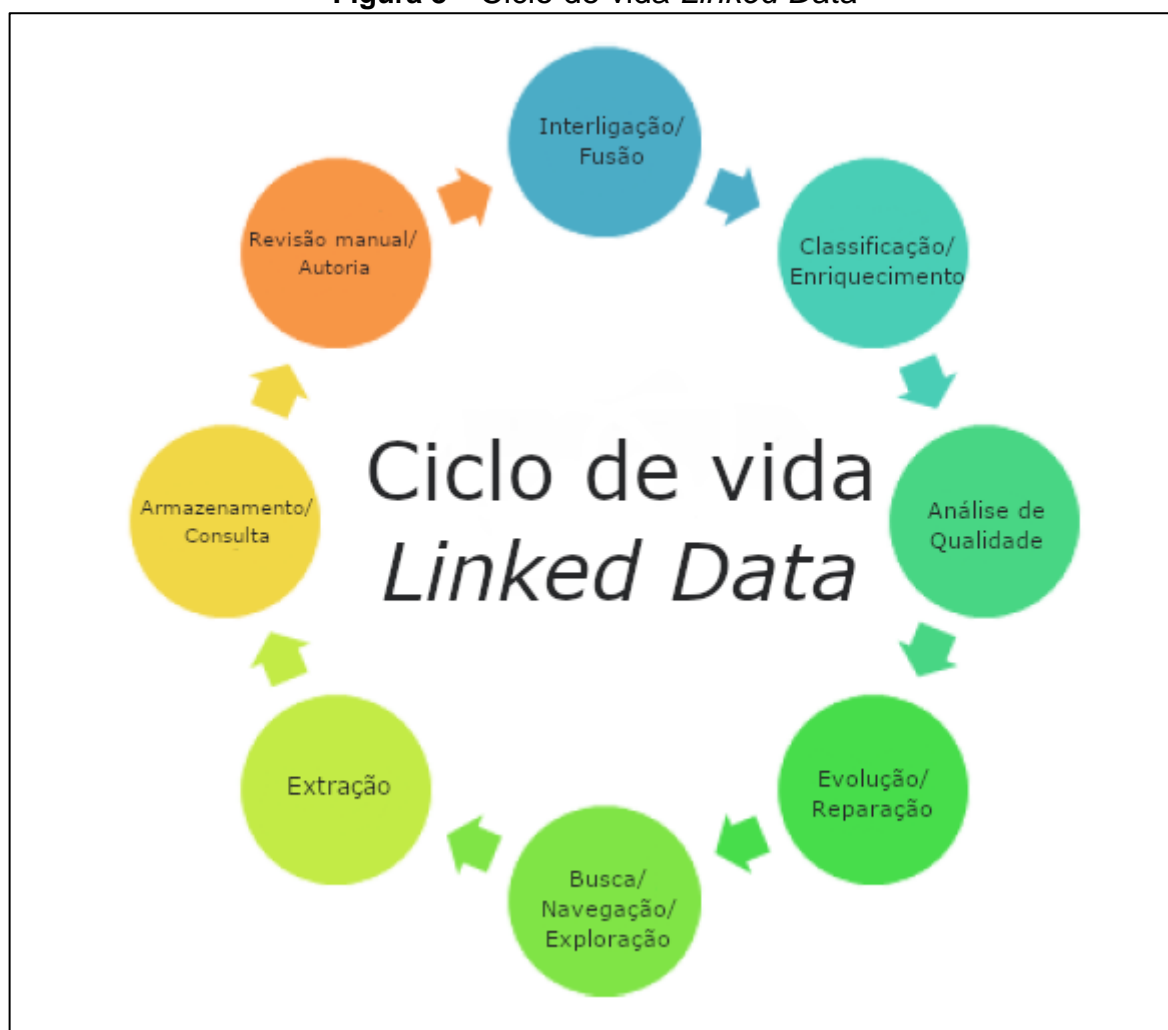
Figura 7 – *Linked Open Data Cloud (LOD Cloud)* de 26 de janeiro de 2021



Fonte: McCrae *et al.* ([2021?]).

Como forma de contribuir na publicação de *Linked Data*, o Instituto de Pesquisa *Agile Knowledge Engineering and Semantic Web (AKSW)* disponibiliza o ciclo de vida *Linked Data (Linked Data Lifecycle)*, sendo “[...] um processo metodológico que contempla as atividades para publicar dados conectados na web.” (RAUTENBERG *et al.*, 2018, p. 73). O ciclo possui oito etapas, conforme pode ser visualizado na Figura 8.

Figura 8 – Ciclo de vida *Linked Data*



Fonte: tradução de Auer *et al.* (2012).

Conforme abordado por Rautenberg *et al.* (2018, p. 74-75, grifos dos autores), as atividades do ciclo de vida *Linked Data* (abordado pelos autores como ciclo de vida de dados conectados) consistem em:

- **Extração:** os dados não estruturados ou estruturados em diferentes formatos ou linguagens (CSV, SQL, respectivamente), ou provenientes de sistemas legados, necessitam ser mapeados para o modelo de dados RDF.
- **Armazenamento/Consulta:** o gerenciamento de dados RDF é realizado por meio do uso de *Triple Stores* como forma de potencializar as tarefas de publicação e consumo de dados.
- **Revisão manual/Autoria:** as tarefas de editoração de dados e a disponibilização destes, segundo os formatos permitidos.

- **Interligação/Fusão:** os dados de uma base são interligados a outros dados, de outras bases, ampliando os contextos de pesquisa e consulta.
- **Classificação/Enriquecimento:** seu objetivo é aumentar a expressividade e a riqueza semântica de uma base de conhecimento em relação a um contexto, representando os dados com ontologias de nível superior ou vocabulários.
- **Análise de Qualidade:** o tratamento dos aspectos de integridade, precisão, consistência e validade de dados é realizado nesta atividade. De forma geral, verificam-se os quesitos de consistência, concisão, compreensibilidade, disponibilidade e proveniência do modelo de dados.
- **Evolução/Reparação:** uma vez encontradas inconsistências nos dados disponibilizados ou no modelo de representação perante requisitos estabelecidos (novos ou não), ações podem ser automatizadas para a correção das não-conformidades.
- **Busca/Navegação/Exploração:** as técnicas de busca, navegação, exploração ou visualização são usadas para manipular dados conectados em diferentes aplicações.

Para cada uma dessas atividades são recomendadas diferentes ferramentas para auxiliar no processo de gestão do ciclo de vida de dados em *Linked Data*, sendo o conjunto delas chamado de *Linked Data Stack* (RAUTENBERG *et al.*, 2018).

Desse modo, para a **Extração** dos dados, têm-se as ferramentas: *OpenLink Virtuoso Sponger*, *DBpedia Spotlight*, *PoolParty Thesaurus Server* (PPT), D2R, R2R, *Apache Stanbol* e *CSVImport* (RAUTENBERG *et al.*, 2018).

Para o **Armazenamento/Consulta** dos dados RDF são utilizados *OpenLink Virtuoso RDF Store*, *SparQLed*, *Sparqlify* e SIREn (RAUTENBERG *et al.*, 2018).

Para a **Revisão manual/Autoria** dos recursos têm-se o *OntoWiki*, *RDFauthor* e *PoolParty Thesaurus Server* (RAUTENBERG *et al.*, 2018).

Para a **Interligação/Fusão** dos dados são adotadas as ferramentas LIMES, *Silk* e SIEVE (RAUTENBERG *et al.*, 2018).

A atividade de **Classificação/Enriquecimento** dos dados em *Linked Data* é adotada a ferramenta *DL-Learner*, enquanto a atividade **Análise de Qualidade** conta com o uso da ferramenta ORE (RAUTENBERG *et al.*, 2018).

Para a atividade de **Evolução/Reparação** de dados utiliza-se o *OpenRefine*. E, por fim, a atividade **Busca/Navegação/Exploração** utiliza-se das ferramentas *Sigma*, *Spatial Semantic Browser*, *CubeViz* e *Facete* (RAUTENBERG *et al.*, 2018).

Desse modo, o ciclo de vida de dados em *Linked Data* conta com atividades que visam orientar a publicação dos dados e proporcionar a sua exploração no contexto da *Web* de dados.

4.1 Tecnologias semânticas

Conforme abordado, os princípios *Linked Data*, bem como a representação da informação no âmbito da *Web*, se apoiam em tecnologias semânticas que visam proporcionar a melhor disponibilização de conteúdo na *Web*. Nesse contexto, visando compreender melhor algumas dessas tecnologias, são apresentados: o 4.1.1 *RDF Schema*, o 4.1.2 *XML Schema* e o 4.1.3 *Simple Knowledge Organization System (SKOS)*.

4.1.1 *RDF Schema*

Definido pelo W3C como uma “extensão semântica” do RDF, o *RDF Schema* é escrito em RDF, sendo um vocabulário destinado à modelagem de dados em RDF (BRICKLEY; GUHA, 2014), de modo que seja possível estruturar os grafos RDF. Nas palavras de Ferreira e Santos (2013, p. 15), trata-se de “[...] uma extensão do RDF, [que] proporciona um meio para a publicação tanto de vocabulários legíveis por humanos quanto de vocabulários legíveis por máquinas.”. Segundo Allemang e Hendler (2008, p. 91, tradução nossa) o *RDF Schema* busca responder:

Quais indivíduos são relacionados entre si e como? Como as propriedades que usamos para definir nossos indivíduos estão relacionadas a outros conjuntos de indivíduos e, de fato, um para o outro? O RDFS fornece uma maneira para um modelador de informações expressar as respostas a esses tipos de perguntas, pois elas dizem respeito a necessidades específicas de modelagem e integração de dados.

De acordo com Shirado e Ramalho (2019, p. 301), tanto o RDF quanto o *RDF Schema* “[...] são ferramentas responsáveis por exprimir significância e promover a interoperabilidade entre metadados e padrões ou formatos de metadados.”.

O RDF *Schema* é formado por classes e propriedades, sendo que “[...] descreve propriedades em termos das classes de recurso às quais se aplicam.” (BRICKLEY; GUHA, 2014, n.p., tradução nossa), o que facilita o processo de adição de propriedades a uma determinada declaração já realizada.

Os recursos se dividem em diferentes classes, sendo que cada uma de suas partes são chamadas de instâncias, podendo ser utilizada a propriedade ‘rdfs:type’ para indicar que um recurso é a instância de uma determinada classe. Duas classes podem ter um mesmo conjunto de instâncias, porém ser diferentes classes. As classes contidas em outras classes são denominadas subclasses, sendo que todas as instâncias de uma subclasse são também instâncias de sua classe – ou seja, da classe superior à subclasse, também chamada de superclasse (BRICKLEY; GUHA, 2014).

As classes principais encontram-se descritas no Quadro 4 abaixo:

Quadro 4 – Classes principais do *Resource Description Framework Schema* (RDF *Schema*)

Classe	Descrição
rdfs:Resource	Tudo que é descrito por RDF é um recurso e são instâncias da classe rdfs:Resource. Todas as outras classes são subclasses dessa classe.
rdfs:Class	Classe dos recursos que são classes RDF.
rdfs:Literal	Classe dos valores literais como cadeias de caracteres (<i>strings</i>) e inteiros (<i>integers</i>).
rdfs:Datatype	Classe de todos os tipos de dados (<i>datatypes</i>).
rdf:langString	Classe das cadeias de caracteres de etiqueta de linguagem (<i>language-tagged string</i>).
rdf:HTML	(Seção não-normativa) Classe dos valores literais HTML.
rdf:XMLLiteral	(Seção não-normativa) Classe dos valores literais XML.
rdf:Property	Classe das propriedades RDF.

Fonte: adaptação e tradução de Brickley e Guha (2014).

O RDF *Schema* conta, ainda, com classes adicionais e propriedades. No Quadro 5 a seguir estão presentes as classes adicionais.

Quadro 5 – Classes adicionais do *Resource Description Framework Schema* (RDF *Schema*)

Tipo de classe	Classe	Descrição
RDF <i>containers</i>	rdfs:Container	(Seção não-normativa) Classe de todos os containers RDF.
	rdf:Bag	(Seção não-normativa) Classe dos containers não-

		ordenados.
	rdf:Seq	(Seção não-normativa) Classe dos containers ordenados.
	rdf:Alt	(Seção não-normativa) Classe dos containers alternativos.
	rdfs:ContainerMembershipProperty	(Seção não-normativa) Classe das propriedades de membros de containers
Coleções RDF (RDF <i>Collections</i>)	rdf:List	(Seção não-normativa) Classe das listas RDF.
Vocabulário de reificação (<i>Reification Vocabulary</i>)	rdf:Statement	(Seção não-normativa) Classe que representa as declarações RDF.

Fonte: adaptação e tradução de Brickley e Guha (2014).

As classes ‘rdfs:Container’, ‘rdf:Bag’, ‘rdf:Seq’, ‘rdf:Alt’ e ‘rdfs:ContainerMembershipProperty’ fazem parte do RDF *containers*, recursos usados para representar coleções abertas – aquelas que “[...] não definem nenhum mecanismo para afirmar que não há mais membros.” (BRICKLEY; GUHA, 2014, n.p., tradução nossa). A classe ‘rdf:List’ pertence ao RDF *Collections*, vocabulário de classes e propriedades que, diferentemente do RDF *containers*, descreve coleções fechadas. Já o ‘rdf:Statement’ faz parte do vocabulário de reificação do RDFS.

Quanto às propriedades do RDF *Schema*, as mesmas são compreendidas como “[...] a relação entre recursos de sujeito e recursos de objeto.” (BRICKLEY; GUHA, 2014, n.p., tradução nossa). Existem ainda as subpropriedades, indicadas por meio da propriedade ‘rdfs:subPropertyOf’, que indica que uma propriedade é hierarquicamente englobada por outra propriedade (BRICKLEY; GUHA, 2014).

As propriedades podem também ser chamadas de instâncias de ‘rdf:Property’. As principais propriedades encontram-se dispostas no Quadro 6:

Quadro 6 – Propriedades principais do *Resource Description Framework Schema* (RDF *Schema*)

Propriedade	Descrição
rdfs:range	Utilizada para declarar que os valores de uma propriedade são instâncias de uma ou mais classes.
rdfs:domain	Utilizada para declarar que qualquer recurso que tenha uma determinada propriedade é uma instância de uma ou mais classes.
rdf:type	Utilizada para declarar que um recurso é uma instância de uma classe.
rdfs:subClassOf	Utilizada para declarar que todas as instâncias de uma classe

	são instâncias de outra classe.
rdfs:subPropertyOf	Utilizada para declarar que todos os recursos relacionados por uma propriedade são também relacionados por outra propriedade.
rdfs:label	Pode ser utilizada para prover uma versão legível por humanos do nome de um recurso.
rdfs:comment	Pode ser utilizada para prover uma versão legível por humanos da descrição de um recurso.

Fonte: adaptação e tradução de Brickley e Guha (2014).

No Quadro 7, são apresentadas as propriedades adicionais do RDF *Schema*.

Quadro 7 – Propriedades adicionais do *Resource Description Framework Schema* (RDF *Schema*)

Tipo de propriedade	Propriedade	Descrição
RDF <i>containers</i>	rdfs:member	(Seção não-normativa) Super-propriedade de todas as propriedades de membros de containers.
Coleções RDF (RDF <i>Collections</i>)	rdf:first	(Seção não-normativa) Utilizada para indicar o primeiro item na lista RDF de sujeito.
	rdf:rest	(Seção não-normativa) Utilizada para indicar o restante dos itens na lista RDF de sujeito depois do primeiro item.
	rdf:nil	(Seção não-normativa) Utilizada para representar uma lista vazia.
Propriedades utilitárias (<i>Utility Properties</i>)	rdfs:seeAlso	Utilizada para fornecer informações adicionais sobre o recurso sujeito.
	rdfs:isDefinedBy	Utilizada para definir o recurso sujeito.
	rdf:value	Propriedade idiomática usada para valores estruturados.
Vocabulário de reificação (<i>Reification Vocabulary</i>)	rdf:subject	(Seção não-normativa) Utilizada para indicar o sujeito das declarações em RDF.
	rdf:predicate	(Seção não-normativa) Utilizada para indicar o predicado das declarações RDF.
	rdf:object	(Seção não-normativa) Utilizada para indicar o objeto das declarações RDF.

Fonte: adaptação e tradução de Brickley e Guha (2014).

Assim como algumas das classes do RDF *Schema*, a propriedade 'rdfs:member' é do tipo *container*, significando que as características das propriedades não são necessariamente atribuídas aos seus membros, pelo princípio

de subpropriedade ou superpropriedade: “Assim como um galinheiro pode ter a propriedade de ser feito de madeira, isso não significa que todas as galinhas que ele contém sejam feitas de madeira [...]” (BRICKLEY; GUHA, 2014, n.p., tradução nossa). Já as propriedades ‘rdf:first’, ‘rdf:rest’ e ‘rdf:nil’ fazem parte do vocabulário RDF *Collections*. No contexto das propriedades, tem-se ainda as propriedades utilitárias ‘rdfs:seeAlso’, ‘rdfs:isDefinedBy’ e ‘rdf:value’, e as propriedades do vocabulário de reificação ‘rdf:subject’, ‘rdf:predicate’ e ‘rdf:object’.

4.1.2 XML Schema

O XML Schema é uma linguagem que busca estabelecer como um documento XML deve ser interpretado, expressando restrições sobre esses documentos (WORLD WIDE WEB CONSORTIUM, 2015a). De acordo com o *World Wide Web Consortium* (2015a, n.p., tradução nossa), o XML Schema pode ser utilizado para:

- fornecer uma lista de elementos e atributos em um vocabulário;
- associar tipos, como número inteiro, string, etc., ou mais especificamente, como *hatsize*, *sock_colour* etc., com valores encontrados nos documentos;
- restringir onde os elementos e atributos podem aparecer e o que pode aparecer dentro desses elementos, como dizer que um título de capítulo ocorre dentro de um capítulo e que um capítulo deve consistir em um título de capítulo seguido por um ou mais parágrafos de texto;
- fornecer documentação legível por humanos e processável por máquina;
- fornecer uma descrição formal de um ou mais documentos.

Nota-se que a verificação com um esquema, chamada de validação, é um processo essencial para a garantia da qualidade de descrição do documento (WORLD WIDE WEB CONSORTIUM, 2015a). Os elementos que compõem o XML Schema encontram-se listados no Quadro 8 a seguir.

Quadro 8 – Elementos do XML Schema

Elementos			
all	complexType	key	Redefine
annotation	documentation	keyref	Restriction
any	element	length	Schema
anyAttribute	enumeration	list	Selector
appinfo	extension	maxInclusive	Sequence

attribute	field	maxLength	simpleContent
attributeGroup	group	minInclusive	simpleType
choice	Import	minLength	Union
complexContent	include	pattern	Unique

Fonte: Fallside e Walmsley (2004).

No Quadro 9 são listados os atributos do XML Schema.

Quadro 9 – Atributos do XML Schema

Atributos		
Abstract	Form	Nillable
attributeFormDefault	itemType	processContents
Base	memberTypes	Ref
Block	maxOccurs	schemaLocation
blockDefault	minOccurs	substitutionGroup
Default	Mixed	targetNamespace
elementFormDefault	name	Type
Final	namespace	xsi:type
finalDefault	xsi:noNamespaceSchemaLocation	Use
Fixed	xsi:nil	Xpath

Fonte: Fallside e Walmsley (2004).

Na Figura 9 a seguir é apresentado um exemplo de uso dos elementos 'complexType', 'name', 'ref' e 'all' e do atributo 'name' do XML Schema.

Figura 9 – Exemplo de uso do XML Schema

```

<xsd:complexType name="PurchaseOrderType">
  <xsd:all>
    <xsd:element name="shipTo" type="USAddress"/>
    <xsd:element name="billTo" type="USAddress"/>
    <xsd:element ref="comment" minOccurs="0"/>
    <xsd:element name="items" type="Items"/>
  </xsd:all>
  <xsd:attribute name="orderDate" type="xsd:date"/>
</xsd:complexType>

```

Fonte: Fallside e Walmsley (2004).

Na Figura 9 o tipo complexo (xsd:complexType) *PurchaseOrderType* é definido, contendo declarações de elementos (xsd:element) e de um atributo (xsd:attribute), sendo que “As declarações não são elas mesmas tipos, mas uma associação entre um nome e as restrições que governam a aparência desse nome em documentos governados pelo esquema associado.” (FALLSIDE; WALMSLEY, 2004, n.p., tradução nossa).

4.1.3 *Simple Knowledge Organization System (SKOS)*

O SKOS “[...] é um modelo de dados comum relacionado ao compartilhamento e à conectividade de sistemas de organização do conhecimento no contexto da web de dados.” (RAUTENBERG *et al.*, 2018, p. 45), de modo a possibilitar a representação padronizada desses sistemas de organização do conhecimento a partir da utilização do RDF, sendo legíveis por máquina e intercambiáveis no contexto da *Web* e *Web Semântica* (ISAAC; PHIPPS; RUBIN, 2009; ISAAC; SUMMERS, 2009; MILES; BECHHOFFER, 2009; WORLD WIDE WEB CONSORTIUM, 2012a; RAMALHO, 2015; 2016). Pode ser compreendido como um modelo de dados, ou um vocabulário RDF, composto por um conjunto definido de conceitos identificados por URIs, objetivando o maior uso e interoperabilidade dos ‘vocabulários conceituais originais’ (ISAAC; SUMMERS, 2009). Ramalho (2015) corrobora com a definição de Isaac e Summers (2009), para o qual “[...] o SKOS não tem a função de substituir vocabulários controlados em seu contexto original de uso, e sim favorecer uma maior reutilização e interoperabilidade entre os vocabulários existentes.”.

Conforme coloca Isaac e Summers (2009, n.p., tradução nossa), o SKOS

[...] fornece um modelo para expressar a estrutura básica e o conteúdo de esquemas de conceitos, como tesouros, esquemas de classificação, listas de cabeçalhos de assuntos, taxonomias, folksonomias e outros tipos semelhantes de vocabulário controlado.

A utilização do SKOS ocorre em nível básico ou avançado, de acordo com a necessidade de descrição dos recursos. Destaca-se que

No SKOS básico, os recursos conceituais (conceitos) são identificados com URIs, rotulados com strings em uma ou mais linguagens naturais, documentados com vários tipos de notas, semanticamente relacionados entre si em hierarquias informais e redes de associação, e agregados em esquemas de conceitos. No SKOS avançado, os recursos conceituais podem ser mapeados através de esquemas de conceito e agrupados em coleções rotuladas ou ordenadas. Os relacionamentos podem ser especificados entre rótulos de conceito. Finalmente, o próprio vocabulário SKOS pode ser estendido para atender às necessidades de comunidades de prática específicas ou combinado com outros vocabulários de modelagem. (ISAAC; SUMMERS, 2009, n.p., tradução nossa).

O vocabulário do SKOS é, como nos vocabulários apresentados anteriormente, formado por classes e propriedades. No Quadro 10 abaixo encontram-se listadas as classes SKOS acompanhadas de suas respectivas descrições.

Quadro 10 – Classes do SKOS

Classe	Definição
skos:Collection	Pertence às Coleções de Conceito (<i>Concept Collections</i>). As coleções são úteis quando um grupo de conceitos compartilha algo em comum, e é conveniente agrupá-las sob um rótulo comum ou onde alguns conceitos podem ser colocados em uma ordem significativa.
skos:Concept	Trata-se da classe de conceitos SKOS. Um conceito SKOS pode ser visto como uma ideia ou noção; uma unidade de pensamento. No entanto, o que constitui uma unidade de pensamento é subjetivo e essa definição deve ser sugestiva e não restritiva.
skos:ConceptScheme	Um esquema de conceito SKOS pode ser visto como uma agregação de um ou mais conceitos SKOS. Os relacionamentos semânticos (<i>links</i>) entre esses conceitos também podem ser vistos como parte de um esquema de conceitos. Essa definição, no entanto, pretende ser mais sugestiva do que restritiva.
skos:OrderedCollection	Pertence às Coleções de Conceito (<i>Concept Collections</i>). Trata-se de uma subclasse de skos:Collection.

Fonte: adaptação e tradução de Miles e Bechhofer (2009).

No Quadro 11 abaixo estão presentes as propriedades SKOS, também acompanhadas de suas definições.

Quadro 11 – Propriedades do SKOS

Propriedade	Definição
skos:altLabel	Pertence aos Rótulos Lexicais (<i>Lexical Labels</i>). Um rótulo lexical é uma sequência de caracteres UNICODE em um determinado idioma natural. Exemplo: “amor romântico” em português ou “れん あ い” em japonês. Trata-se de uma subpropriedade de rdfs:label.
skos:broadMatch	Pertence às Propriedades de Mapeamento (<i>Mapping Properties</i>). Essas propriedades são usadas para indicar <i>links</i> de mapeamento (alinhamento) entre conceitos SKOS em diferentes esquemas de conceitos, em que os <i>links</i> são inerentes ao significado dos conceitos vinculados. Trata-se de uma subpropriedade de skos:mappingRelation e skos:broader.
skos:broader	Pertence às Relações Semânticas (<i>Semantic Relations</i>). As relações semânticas do SKOS são <i>links</i> entre os conceitos do SKOS, nos quais o <i>link</i> é inerente ao significado dos conceitos vinculados. Trata-se de uma subpropriedade de skos:broaderTransitive.
skos:broaderTransitive	Pertence às Relações Semânticas (<i>Semantic Relations</i>).

Propriedade	Definição
	Trata-se de uma subpropriedade de skos:semanticRelation.
skos:changeNote	Pertence às Propriedades de Documentação (<i>Documentation Properties</i>). As notas são usadas para fornecer informações relacionadas aos conceitos do SKOS. Trata-se de uma subpropriedade de skos:note.
skos:closeMatch	Pertence às Propriedades de Mapeamento (<i>Mapping Properties</i>). Trata-se de uma subpropriedade de skos:mappingRelation.
skos:definition	Pertence às Propriedades de Documentação (<i>Documentation Properties</i>). Trata-se de uma subpropriedade de skos:note.
skos:editorialNote	Pertence às Propriedades de Documentação (<i>Documentation Properties</i>). Trata-se de uma subpropriedade de skos:note.
skos:exactMatch	Pertence às Propriedades de Mapeamento (<i>Mapping Properties</i>). Trata-se de uma subpropriedade de skos:closeMatch.
skos:Example	Pertence às Propriedades de Documentação (<i>Documentation Properties</i>). Trata-se de uma subpropriedade de skos:note.
skos:hasTopConcept	Pertence aos Esquemas de Conceito (<i>Concept Schemes</i>).
skos:hiddenLabel	Pertence aos Rótulos Lexicais (<i>Lexical Labels</i>). Trata-se de uma subpropriedade de rdfs:label.
skos:historyNote	Pertence às Propriedades de Documentação (<i>Documentation Properties</i>). Trata-se de uma subpropriedade de skos:note.
skos:inScheme	Pertence aos Esquemas de Conceito (<i>Concept Schemes</i>).
skos:mappingRelation	Pertence às Propriedades de Mapeamento (<i>Mapping Properties</i>). Trata-se de uma subpropriedade de skos:semanticRelation.
skos:member	Pertence às Coleções de Conceito (<i>Concept Collections</i>).
skos:memberList	Pertence às Coleções de Conceito (<i>Concept Collections</i>). Para qualquer recurso, cada item da lista fornecido como o valor da propriedade skos:memberList também é um valor da propriedade skos:member.
skos:narrowMatch	Pertence às Propriedades de Mapeamento (<i>Mapping Properties</i>). Trata-se de uma subpropriedade de skos:mappingRelation e de skos:narrower.
skos:narrower	Pertence às Relações Semânticas (<i>Semantic Relations</i>). Trata-se de uma subpropriedade de skos:narrowerTransitive.
skos:narrowerTransitive	Pertence às Relações Semânticas (<i>Semantic Relations</i>). Trata-se de uma subpropriedade de skos:semanticRelation.
skos:notation	Utilizado para notações. Uma notação é uma sequência de caracteres que não é reconhecida como uma palavra em qualquer linguagem natural, como "T58.5" ou "303.4833", usada para identificar exclusivamente um conceito dentro do escopo de um determinado esquema de conceito.
skos:note	Pertence às Propriedades de Documentação (<i>Documentation Properties</i>).
skos:prefLabel	Pertence aos Rótulos Lexicais (<i>Lexical Labels</i>). Trata-se de subpropriedade de rdfs:label.
skos:related	Pertence às Relações Semânticas (<i>Semantic Relations</i>). Trata-se de uma subpropriedade de skos:semanticRelation.
skos:relatedMatch	Pertence às Propriedades de Mapeamento (<i>Mapping Properties</i>). Trata-se de subpropriedade de skos:mappingRelation e de skos:related.
skos:scopeNote	Pertence às Propriedades de Documentação (<i>Documentation</i>

Propriedade	Definição
	<i>Properties</i>). Trata-se de subpropriedade de skos:note.
skos:semanticRelation	Pertence às Relações Semânticas (<i>Semantic Relations</i>).
skos:topConceptOf	Pertence aos Esquemas de Conceito (<i>Concept Schemes</i>). Trata-se de uma subpropriedade de skos:inScheme.

Fonte: adaptação e tradução de Miles e Bechhofer (2009).

A Figura 10 apresenta exemplo de uso de classe e propriedade do SKOS. Nela consta que uma coleção (skos:Collection) chamada *MyCollection* possui três membros (skos:member): X, Y e Z.

Figura 10 – Exemplo de uso de classe do SKOS

```
<MyCollection> rdf:type skos:Collection ;
  skos:member <X> , <Y> , <Z> .
```

Fonte: Miles e Bechhofer (2009).

Ressalta-se que existem classes e propriedades adicionais fornecidas pelo SKOS *eXtension for Labels* (SKOS-XL), sendo de caráter opcional (MILES; BECHHOFER, 2009).

Desse modo, ao utilizar o SKOS é possível realizar a identificação de conceitos pelo uso de URIs, bem como atribuir variadas descrições a eles e estabelecer relações entre os conceitos (MILES; BECHHOFER, 2009).

4.2 Princípios FAIR

Os princípios FAIR foram publicados em 2016, sendo seu objetivo permitir que os dados sejam encontráveis (*findable*), acessíveis (*accessible*), interoperáveis (*interoperable*) e reutilizáveis (*reusable*) (FORCE11, c2017). De acordo com Wilkinson *et al.* (2016), os princípios foram elaborados para auxiliar os produtores e publicadores de dados na gestão dos dados de pesquisa, de modo a proporcionar a descoberta e o reuso da informação.

Nota-se que os princípios são orientações – não sendo um padrão nem uma especificação – visando guiar o processo de publicação e avaliação dos recursos de pesquisa, de forma que os mesmos sejam, ou estejam, encontráveis, acessíveis, interoperáveis e reutilizáveis (WILKINSON *et al.*, 2016). Destaca-se que em âmbito internacional já existe um consenso quanto à recomendação para a publicação dos dados segundo os princípios FAIR (LAENDER *et al.*, 2020).

Segundo Wilkinson *et al.* (2016), os princípios FAIR buscam permitir que os dados sejam semanticamente compreensíveis tanto por humanos quanto por máquinas.

O estado ideal – onde as máquinas "entendem" completamente e podem autonomamente e corretamente operar um objeto digital – raramente pode ser alcançado. No entanto, os princípios FAIR fornecem 'etapas ao longo do caminho' em direção à acionabilidade da máquina; a adoção, no todo ou em parte, dos princípios FAIR, leva o recurso ao longo do continuum em direção a esse estado ideal. (WILKINSON *et al.*, 2016, p. 3 e 4, tradução nossa).

Para os autores, essa acionabilidade da máquina (*machine-actionability*) a ser alcançada corresponde à capacidade da máquina realizar ações autônomas, sem antes ter encontrado o objeto, a partir da descrição detalhada e completa dos recursos digitais, sendo: a identificação do tipo do objeto; a utilidade do recurso ao contexto de determinado agente; identificar se o recurso é utilizável, considerando licenças, consentimentos e restrições; e tomada de ações apropriadas (WILKINSON *et al.*, 2016). Ou como a GO FAIR ([2019?]), coloca, é “[...] a capacidade dos sistemas computacionais de encontrar, acessar, interoperar e reutilizar dados com nenhuma ou mínima intervenção humana.”.

Dentre os princípios FAIR, observa-se que não se encontra presente uma perspectiva quanto à qualidade dos dados, sendo que esta é abordada pelos ciclos de vida dos dados (DIAS; ANJOS; RODRIGUES, 2019), conforme discutido anteriormente.

No Quadro 12 a seguir encontram-se dispostos os princípios FAIR:

Quadro 12 – Princípios FAIR

Encontrabilidade	F1. Aos (meta)dados são atribuídos um identificador globalmente exclusivo e persistente.
	F2. Os dados precisam ser descritos com metadados ricos.
	F3. Os metadados incluem clara e explicitamente o identificador dos dados que descrevem.
	F4. Os (meta)dados são registrados ou indexados em um recurso pesquisável.
Acessibilidade	A1. Os (meta)dados são recuperáveis pelo seu identificador usando um protocolo de comunicação padronizado.
	A1.1. O protocolo é aberto, gratuito e universalmente implementável.
	A1.2. O protocolo permite um procedimento de autenticação e autorização, quando necessário.
	A2. Os metadados são acessíveis, mesmo quando os dados não estão

	mais disponíveis.
Interoperabilidade	I1. Os (meta)dados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.
	I2. Os (meta)dados utilizam vocabulários e/ou ontologias que seguem os princípios FAIR.
	I3. Os (meta)dados incluem referências qualificadas a outros (meta)dados.
Reutilização	R1. Os (meta)dados são ricamente descritos com uma pluralidade de atributos precisos e relevantes.
	R1.1. Os (meta)dados são liberados com uma licença de uso de dados clara e acessível.
	R1.2. Os (meta)dados estão associados com a proveniência detalhada.
	R1.3. Os (meta)dados cumprem os padrões comunitários relevantes para o domínio.

Fonte: adaptação e tradução de GO FAIR ([2019?]).

Nota-se que o uso do termo “(meta)dados” implica que o princípio se aplica tanto aos dados quanto aos metadados. Desse modo, para os dados serem encontráveis orienta-se que: os dados e metadados possuam um identificador único, como um URL (F1); os dados sejam descritos com metadados ricos, de modo abrangente (F2); os metadados possuam identificadores dos dados sendo descritos (F3); os dados e metadados possam ser pesquisáveis, para que de fato sejam encontráveis (F4) (GO FAIR, [2019?]).

Para os dados serem acessíveis, orienta-se que: seja utilizado um protocolo de comunicação padronizado (A1), sendo o protocolo aberto, gratuito e globalmente implementável (A1.1) e que permita que a máquina realize a autenticação e autorização para o acesso aos dados, caso necessário (A1.2); os metadados estejam disponíveis, mesmo quando os dados já não estão (A2) (GO FAIR, [2019?]).

Pensando-se na interoperabilidade, orienta-se que: os dados e metadados utilizem linguagens compreensíveis, tanto para humano quanto para máquinas, para representação do conhecimento, como vocabulários controlados, ontologias e tesouros amplamente usados (I1); os dados e metadados usem vocabulários que seguem os princípios FAIR, sendo encontráveis e acessíveis (I2); os dados e metadados estabeleçam referências bem especificadas a outros dados e metadados, ou seja, explicitando a relação existente entre os mesmos (I3) (GO FAIR, [2019?]).

Por fim, para os dados serem reutilizáveis, os princípios FAIR orientam que: os dados e metadados sejam descritos ricamente, de modo que o usuário possua informações suficientes para decidir se os dados e metadados são relevantes para o

seu contexto (R1) – esse princípio se relaciona com o F2, porém com foco no usuário; os dados e metadados adotem licença clara e acessível, de modo a promover a interoperabilidade legal dos mesmos (R1.1); os dados e metadados sejam acompanhados de informações de proveniência detalhadas, como a origem e por quem foram gerados ou coletados, se já fora publicado anteriormente, dentre outros (R1.2); os dados e metadados sigam os padrões da comunidade em que se inserem, facilitando o seu reuso (R1.3) (GO FAIR, [2019?]).

É possível observar que os princípios FAIR reforçam o papel da descrição dos dados de pesquisa, uma vez que todos abordam os metadados, indicando a imprescindibilidade deles quando se deseja uma descoberta e recuperação ótima dos dados. Conforme Simionato (2017, n.p.) aponta, “[...] os metadados destacam-se pela importância no ciclo de vida dos dados científicos, viabilizando a sua disponibilização e reuso dos dados [...]”.

Ademais, autores como Fonseca e Monteiro (2019), Henning *et al.* (2019), Monteiro e Sant’Ana (2019), dentre outros, também buscam explorar os princípios FAIR no contexto de dados abertos (FONSECA; MONTEIRO, 2019; HENNING *et al.*, 2019; MONTEIRO; SANT’ANA, 2019) e de dados de pesquisa (HENNING *et al.*, 2019; MONTEIRO; SANT’ANA, 2019), a fim de melhor compreender e contribuir à gestão dos dados.

4.3 Melhores práticas para publicação de dados na Web

Pensando no melhor desenvolvimento da *Web*, o W3C, sob edição de Lóscio, Burle e Calegari (2017) introduz 35 melhores práticas (MPs) para a publicação e uso de dados no ambiente.

Nos casos em que os dados são usados de alguma maneira, seja pelo originador dos dados ou por uma parte externa, esse uso também deve ser descoberto e os esforços do editor de dados reconhecidos. Em resumo, seguir essas práticas recomendadas facilitará a interação entre editores e consumidores.

No Quadro 13 a seguir encontram-se as melhores práticas organizadas de acordo com suas categorias, acompanhadas de suas descrições e benefícios, conforme o Lóscio, Burle e Calegari (2017).

Quadro 13 – Melhores práticas para publicação de dados na Web

Categoria	Melhor prática	Descrição	Benefício
Metadados	1. Fornecer metadados	Fornecer metadados para usuários humanos e aplicativos de computador.	Reuso, compreensão, descoberta e processabilidade.
	2. Fornecer metadados descritivos	Fornecer metadados que descrevam as características gerais de conjuntos de dados e distribuições ¹⁰ .	Reuso, compreensão e descoberta.
	3. Fornecer metadados estruturais	Forneça metadados que descrevam o esquema e a estrutura interna de uma distribuição.	Reuso, compreensão e processabilidade.
Licenças de dados	4. Fornecer informações de licença de dados	Fornecer um <i>link</i> ou cópia do contrato de licença que controla o uso dos dados.	Reuso e confiabilidade.
Proveniência de dados	5. Fornecer informações de proveniência de dados	Fornecer informações completas sobre as origens dos dados e quaisquer alterações feitas.	Reuso, compreensão e confiabilidade.
Qualidade de dados	6. Fornecer informações sobre a qualidade dos dados	Fornecer informações sobre a qualidade e adequação dos dados para fins específicos.	Reuso e confiabilidade.
Versionamento de dados	7. Fornecer um indicador de versão	Atribuir e indicar um número ou data da versão para cada conjunto de dados.	Reuso e confiabilidade.
	8. Fornecer histórico de versões	Fornecer um histórico completo da versão que explique as alterações feitas em cada versão.	Reuso e confiabilidade.
Identificadores de dados	9. Usar URIs persistentes como identificadores de conjuntos de dados	Identificar cada conjunto de dados por um URI persistente cuidadosamente escolhido.	Reuso, ligação, descoberta e interoperabilidade.
	10. Usar URIs persistentes como identificadores dentro dos conjuntos de dados	Reutilizar os URIs de outras pessoas como identificadores dentro dos conjuntos de dados, sempre que possível.	Reuso, ligação, descoberta e interoperabilidade.
	11. Atribuir URIs para versões e séries de	Atribuir URIs para versões individuais de	Reuso, descoberta e confiabilidade.

¹⁰ “Uma distribuição representa uma forma específica disponível de um conjunto de dados. Cada conjunto de dados pode estar disponível em diferentes formas; esses formulários podem representar diferentes formatos do conjunto de dados ou diferentes endpoints. Exemplos de distribuições incluem um arquivo CSV para download, uma API ou um feed RSS.” (LÓSCIO; BURLE; CALEGARI, 2017, n.p., tradução nossa).

Categoria	Melhor prática	Descrição	Benefício
	conjuntos de dados	conjuntos de dados, bem como para a série geral.	
Formatos de dados	12. Use formatos de dados padronizados legíveis por máquina	Disponibilizar os dados em um formato de dados padronizado, legível por máquina, adequado para o uso pretendido ou potencial.	Reuso e processabilidade.
	13. Usar representações de dados neutras em termos de idioma e local (<i>locale</i> ¹¹ - <i>neutral</i>)	Usar estruturas de dados e valores <i>locale-neutral</i> ou, quando isso não for possível, fornecer metadados sobre o <i>locale</i> usado por valores de dados.	Reuso e compreensão.
	14. Fornecer dados em múltiplos formatos	Disponibilizar dados em múltiplos formatos quando mais de um formato for adequado ao uso pretendido ou potencial.	Reuso e processabilidade.
Vocabulários de dados	15. Reutilizar vocabulários, preferencialmente padronizados	Usar termos de vocabulários compartilhados, de preferência padronizados, para codificar dados e metadados.	Reuso, processabilidade, compreensão, confiabilidade e interoperabilidade.
	16. Escolher o nível de formalização certo	Optar por um nível de semântica formal que se encaixe nos dados e nos aplicativos mais prováveis.	Reuso, compreensão e interoperabilidade.
Acesso aos dados	17. Fornecer <i>download</i> em massa	Permitir que os consumidores recuperem o conjunto de dados completo com uma única solicitação.	Reuso e acesso.
	18. Fornecer subconjuntos para grandes conjuntos de dados	Se seu conjunto de dados for grande, permitir que usuários e aplicativos trabalhem prontamente com subconjuntos úteis de seus dados.	Reuso, ligação, acesso e processabilidade.

¹¹ “Uma coleção de preferências internacionais, geralmente relacionadas a um idioma e uma região geográfica exigida por uma (certa categoria) de usuários. Geralmente, eles são identificados por um identificador abreviado ou token, como uma etiqueta de idioma, que é passada do ambiente para vários processos para obter um comportamento culturalmente afetado.” (LÓSCIO; BURLE; CALEGARI, 2017, n.p., tradução nossa).

Categoria	Melhor prática	Descrição	Benefício
	19. Usar a negociação de conteúdo ¹² para veicular dados disponíveis em múltiplos formatos	Usar a negociação de conteúdo, além das extensões de arquivo, para veicular dados disponíveis em múltiplos formatos.	Reuso e acesso.
	20. Fornecer acesso em tempo real	Quando os dados são produzidos em tempo real, disponibilizá-los na <i>Web</i> em tempo real ou quase em tempo real.	Reuso e acesso.
	21. Fornecer dados atualizados	Disponibilizar os dados de maneira atualizada e tornar a frequência da atualização explícita.	Reuso e acesso.
	22. Fornecer uma explicação para dados que não estão disponíveis	Para dados que não estão disponíveis, fornecer uma explicação sobre como os dados podem ser acessados e quem pode acessá-los.	Reuso e confiabilidade.
	23. Disponibilizar dados por meio de uma API	Oferecer uma API para veicular dados, se você tiver os recursos para fazê-lo.	Reuso, processabilidade, interoperabilidade e acesso.
	24. Usar os padrões da <i>Web</i> como base das APIs	Ao projetar APIs, usar um estilo arquitetural baseado nas tecnologias da própria <i>Web</i> .	Reuso, ligação, interoperabilidade, descoberta, acesso e processabilidade.
	25. Fornecer documentação completa para sua API	Fornecer informações completas na <i>Web</i> sobre sua API. Atualizar a documentação ao adicionar recursos ou fazer alterações.	Reuso e confiabilidade.
	26. Evitar interromper alterações na sua API	Evitar alterações em sua API que quebrem o código do cliente e comunicar quaisquer alterações em sua API aos desenvolvedores quando a evolução ocorrer.	Confiabilidade e interoperabilidade.
Preservação de dados	27. Preservar identificadores	Ao remover dados da <i>Web</i> , preservar o identificador e fornecer informações sobre o recurso arquivado.	Reuso e confiabilidade.

¹² “A negociação de conteúdo é um ponto de flexibilidade na arquitetura da web [...]”. “Um uso é conduzir um novo formato de dados para um mundo que inicialmente não o aceita. Ele permite que os sistemas que aceitam o novo formato o anunciem e participem da nova tecnologia.” (BERNERS-LEE, 2009, n.p., tradução nossa).

Categoria	Melhor prática	Descrição	Benefício
	28. Avaliar a cobertura do conjunto de dados	Avaliar a cobertura de um conjunto de dados antes de sua preservação.	Reuso e confiabilidade.
<i>Feedback</i>	29. Reunir feedback dos consumidores de dados	Fornecer um meio facilmente detectável para os consumidores oferecerem feedback.	Reuso, compreensão e confiabilidade.
	30. Disponibilizar <i>feedback</i>	Disponibilize publicamente os <i>feedbacks</i> dos consumidores sobre conjuntos de dados e distribuições.	Reuso e confiabilidade.
Enriquecimento de dados	31. Enriquecer dados gerando novos dados	Enriquecer seus dados gerando novos dados, ao fazer isso aumentará seu valor.	Reuso, compreensão, confiabilidade e processabilidade.
	32. Fornecer apresentações complementares	Enriquecer os dados apresentando-os de maneiras complementares e imediatamente informativas, como visualizações, tabelas, aplicações <i>Web</i> ou resumos.	Reuso, compreensão, acesso e confiabilidade.
Republicação	33. Fornecer <i>feedback</i> ao publicador original	Informar o publicador original quando você estiver reutilizando seus dados. Se você encontrar um erro ou tiver sugestões ou elogios, informá-los.	Reuso, interoperabilidade e confiabilidade.
	34. Seguir os termos de licenciamento	Encontrar e seguir os requisitos de licenciamento do publicador original do conjunto de dados.	Reuso e confiabilidade.
	35. Citar a publicação original	Reconhecer a fonte dos seus dados em metadados. Se você fornecer uma interface do usuário, incluir a citação visivelmente na interface.	Reuso, descoberta e confiabilidade.

Fonte: adaptação e tradução de Lóscio, Burle e Calegari (2017).

Os diferentes benefícios listados na quarta coluna do Quadro 13 correspondem a quais melhorias cada uma das melhores práticas promove aos

dados sendo publicados na *Web*. O significado de cada benefício encontra-se listado abaixo:

- **Compreensão:** os seres humanos terão uma melhor compreensão sobre a estrutura de dados, o significado dos dados, os metadados e a natureza do conjunto de dados.
- **Processabilidade:** as máquinas poderão processar e manipular automaticamente os dados em um conjunto de dados.
- **Descoberta:** as máquinas poderão descobrir automaticamente um conjunto de dados ou dados dentro de um conjunto de dados.
- **Reuso:** as chances de reutilização de conjuntos de dados por diferentes grupos de consumidores de dados aumentarão.
- **Confiabilidade:** a confiança que os consumidores têm no conjunto de dados aumentará.
- **Ligação:** será possível criar ligações entre recursos de dados (conjuntos de dados e itens de dados).
- **Acesso:** humanos e máquinas poderão acessar dados atualizados de várias formas.
- **Interoperabilidade:** será mais fácil chegar a um consenso entre os editores de dados e os consumidores. (LÓSCIO; BURLE; CALEGARI, 2017, n.p., tradução e grifo nosso).

Observa-se que as melhores práticas, além de promover e guiar a publicação de dados na *Web*, recomendam a utilização de *Linked Data* (LÓSCIO; BURLE; CALEGARI, 2017), como na recomendação de uso de URIs como identificadores (MP 10) e na reutilização de vocabulários (MP 15).

Silva, Santarém Segundo e Silva (2018) apresentam um alinhamento entre essas melhores práticas e os princípios FAIR. Os autores puderam identificar que existe uma convergência entre as melhores práticas para publicação de dados na *Web* e as tecnologias por elas indicadas e os princípios FAIR, sendo que cerca de 80% desses se encontram contemplados pelas melhores práticas (SILVA; SANTARÉM SEGUNDO; SILVA, 2018).

De acordo com Silva, Santarém Segundo e Silva (2018) “[...] as principais categorias das MPs que atendem aos princípios de FAIR são metadados, licença, proveniência, identificadores persistentes, formatos de dados, vocabulários e acesso a dados.”. No entanto, as demais categorias, apesar de não contemplarem por completo os princípios FAIR, não deixam de serem úteis e agregarem valor à publicação de dados de pesquisa (SILVA; SANTARÉM SEGUNDO; SILVA, 2018).

5 DATA CATALOG VOCABULARY (DCAT)

Sendo os dados recursos informacionais de laboriosa manipulação (GRAY *et al.*, 2002), considera-se a necessidade de uniformização dos termos, sendo utilizados os metadados para a representação dos dados publicados.

Rocha, Sales e Sayão (2017) abordam que os metadados buscam dar sentido aos dados de pesquisa, auxiliando em sua compreensão e interpretação. De acordo com os autores, “Os metadados possuem papel central nas pesquisas científicas: são eles que dão sentido ao conjunto de dados coletados em uma determinada pesquisa, deixando aos pesquisadores o encargo de interpretá-los.” (ROCHA; SAYÃO; SALES, 2017, p. 194). De modo semelhante, Simionato (2017) diz que os metadados ajudam a encontrar e interpretar os dados de pesquisa, sendo que “[...] os padrões de metadados auxiliam no controle de atributos e valores em sistemas informacionais, como os repositórios de dados.” (SIMIONATO, 2017, n.p.).

Por meio do emprego de vocabulários é possível controlar o uso dos termos empregados nos campos de metadados, garantindo a padronização dos mesmos e, por conseguinte, a melhor recuperação dos dados. No contexto da *Web Semântica*, o *World Wide Web Consortium* (c2015c, n.p., tradução nossa) coloca que

O papel dos vocabulários [...] é ajudar a integração de dados quando, por exemplo, podem existir ambiguidades nos termos usados nos diferentes conjuntos de dados, ou quando um pouco de conhecimento extra pode levar à descoberta de novos relacionamentos.

Segundo Pattuelli (2012, p. 559, tradução nossa) os vocabulários são essenciais no contexto dos princípios *Linked Data*, uma vez que facilitam “[...] a integração e o reuso de conteúdo devido a sua capacidade de reduzir a ambiguidade semântica e apoiar efetivamente a interligação entre diferentes conjuntos de dados.”. Rozsa, Dutra e Nhacuongue (2017) também abordam os vocabulários no contexto do *Linked Data*, afirmando que

Os vocabulários permitem a definição de conceitos, propriedades e relações entre diferentes conceitos e são utilizados na estruturação e contextualização dos dados na *Web de Dados* para que possam ser compreendidos e processados automaticamente. (ROZSA; DUTRA; NHACUONGUE, 2017, p. 35).

A aplicação de vocabulários promove, então, a otimização da recuperação da informação pelos usuários e aumenta a eficácia nas atividades envolvidas na gestão e organização dos dados pelos profissionais da informação.

Os vocabulários voltados à padronização dos metadados são, no âmbito da Ciência da Informação, conhecidos como padrões de metadados. Alves (2005, p. 115) os define como “[...] estruturas padronizadas para a representação do conteúdo informacional que será representado pelo conjunto de [...] metadados.”, de modo a favorecer a recuperação e a interoperabilidade dos recursos informacionais.

Gilliland (2016, n.p., tradução nossa) estabelece as seguintes tipologias de padrões (Quadro 14):

Quadro 14 – Tipos de padrões de dados

Tipo	Definição	Exemplo
Padrões de estrutura de dados	Termos que indicam a categoria do dado do recurso a ser preenchido. São os conjuntos e esquemas de elementos de metadados	Dublin Core Metadata Element Set (DCMES); formato Machine-Readable Cataloging (MARC)
Padrões de valores de dados	Termos que preenchem a estrutura de dados. São os vocabulários controlados e tesauros	Library of Congress Subject Headings (LCSH); Getty Art & Architecture Thesaurus
Padrões de conteúdo de dados	Servem de guia para o formato e sintaxe dos valores de dados. São as regras de catalogação e códigos	Anglo-American Cataloguing Rules (AACR); Resource Description and Access (RDA); International Standard Bibliographic Description (ISBD)
Padrões de intercâmbio técnico/formato de dados	São os padrões de metadados expressos em formato legível por máquina	Resource Description Framework (RDF); MARC21

Fonte: adaptação de Gilliland (2016).

No contexto da presente pesquisa são abordados os vocabulários que visam padronizar os elementos de metadados, termos que funcionam como categorias para o preenchimento dos dados. De acordo com as tipologias abordadas por Gilliland (2016), correspondem aos padrões de estrutura de dados.

Dessa forma, aborda-se o DCAT, um vocabulário que foi desenvolvido inicialmente pelo *Deri Enterprise Research Institute* (DERI), tendo sido aperfeiçoado pelo *eGov Interest Group* e padronizado pelo grupo de trabalho *Government Linked Data* (GLD) em 2014. A sua segunda versão atualmente vigente foi desenvolvida

pelo grupo de trabalho *Dataset Exchange*, publicada pelo W3C (ALBERTONI *et al.*, 2020).

O vocabulário busca guiar a publicação de catálogos de dados na *Web*, sendo que foi inicialmente elaborado para catálogos de dados governamentais, mas pode ser aplicado a outros contextos (ALBERTONI *et al.*, 2020). Dessa forma, “O DCAT permite que um publicador descreva conjuntos de dados e serviços de dados em um catálogo usando um modelo e vocabulário padrões que facilita o consumo e a agregação de metadados de múltiplos catálogos.” (ALBERTONI *et al.*, 2020, n.p., tradução nossa). Observa-se também que o DCAT possibilita a descrição de relacionamentos entre os conjuntos de dados e entre os conjuntos de dados e outros recursos (ALBERTONI *et al.*, 2020).

Sendo assim, é possível: melhorar a descoberta de conjuntos de dados e serviços de dados; promover a descentralização da publicação de catálogos de dados e a busca federada em múltiplos catálogos (ALBERTONI *et al.*, 2020).

De acordo com Albertoni *et al.* (2020, n.p., tradução nossa), um catálogo de dados se encontra em conformidade com o DCAT se:

- O acesso aos dados é organizado em conjuntos de dados, distribuições e serviços de dados.
- Uma descrição RDF do próprio catálogo, os recursos catalogados correspondentes e distribuições estão disponíveis (mas a escolha da sintaxe RDF, protocolo de acesso e política de acesso não é exigida por esta especificação).
- O conteúdo de todos os campos de metadados mantidos no catálogo e que contêm dados sobre o próprio catálogo, os recursos catalogados correspondentes e distribuições estão incluídos nesta descrição RDF e são expressas usando as classes e propriedades apropriadas do DCAT, exceto onde não existe essa classe ou propriedade.
- Todas as classes e propriedades definidas no DCAT são usadas de maneira consistente com a semântica declarada nesta especificação.

O Albertoni *et al.* (2020, n.p., tradução nossa) define que “O vocabulário DCAT é uma ontologia OWL2, formalizado a partir do RDF *Schema*”. O OWL2 Web Ontology Language (OWL2) é uma linguagem desenvolvida para o contexto da *Web*

Semântica, sendo que as ontologias OWL2 “[...] fornecem classes, propriedades, indivíduos e valores de dados e são armazenadas como documentos da Web Semântica” (WORLD WIDE WEB CONSORTIUM, 2012b, n.p., tradução nossa). As informações em DCAT podem ser apresentadas em diferentes formas, como RDF, HTML-RDFa, RDF/XML, N3, Turtle, JSON-LD, dentre outros (ALBERTONI *et al.*, 2020).

O DCAT faz uso de diferentes vocabulários para complementar a descrição dos dados, como o DC *Terms*, FOAF Vocabulary, *vCard Ontology* e PROV-O, permitindo a adoção de outros vocabulários em conjunto, como o Void (ALBERTONI *et al.*, 2020). Nos Quadros 15 e 16 a seguir encontra-se descritos, respectivamente, as classes e as propriedades que compõem o DCAT.

Quadro 15 – Classes do *Data Catalog Vocabulary* (DCAT)

Classe	Descrição	Superclasse(s)
dcat:Catalog	Uma coleção de metadados com curadoria sobre recursos (por exemplo, conjuntos de dados e serviços de dados no contexto de um catálogo de dados).	dcat:Dataset
dcat:Resource	Recurso publicado ou curado por um único agente.	-
dcat:CatalogRecord	Um registro em um catálogo, descrevendo o registro de um único recurso (dcat:Resource).	-
dcat:Dataset	Uma coleção de dados, publicada ou com curadoria feita por um único agente, e disponível para acessar ou baixar em uma ou mais representações.	dcat:Resource
dcat:Distribution	Uma representação específica de um conjunto de dados. Um conjunto de dados pode estar disponível em várias serializações que podem diferir de várias maneiras, incluindo linguagem natural, tipo de mídia ou formato, organização esquemática, resolução temporal e espacial, nível de detalhes ou perfis (que podem especificar um ou todos os itens acima).	-
dcat:DataService	Uma coleção de operações que fornece acesso a um ou mais conjuntos de dados ou funções de processamento de dados.	dcat:Resource dctype:Service
skos:ConceptScheme	Um sistema de organização do conhecimento (SOC) usado para representar temas/categorias de conjuntos de dados no catálogo.	-
skos:Concept	Uma categoria ou um tema usado para descrever conjuntos de dados no catálogo.	-

Classe	Descrição	Superclasse(s)
foaf:Person	Para pessoas.	
foaf:Organization	Para agências do governo e outras entidades.	
dcat:Relationship	Uma classe de associação para anexar informações adicionais a um relacionamento entre Recursos DCAT	prov:EntityInfluence
dcat:Role	Um papel (dcat:Role) é a função de um recurso ou agente em relação a outro recurso, no contexto de atribuição de recursos ou relacionamentos de recursos.	skos:Concept
dct:PeriodOfTime	Um intervalo de tempo que é nomeado ou definido por seu início e fim.	-
dct:Location	Uma região espacial ou local nomeado.	-

Fonte: Adaptação e tradução de Albertoni *et al.* (2020).

Destaca-se que das classes listadas, o DCAT possui foco em seis classes principais, sendo: 'dcat:Catalog', 'dcat:Resource', 'dcat:Dataset', 'dcat:Distribution', 'dcat:DataService' e 'dcat:CatalogRecord' (ALBERTONI *et al.*, 2020). No Quadro 16 a seguir são listadas as propriedades do DCAT.

Quadro 16 – Propriedades do *Data Catalog Vocabulary* (DCAT)

Propriedade	Descrição	Classe
foaf:homepage	Uma página inicial do catálogo (um documento público da Web geralmente disponível em HTML).	dcat:Catalog
dcat:themeTaxonomy	Um sistema de organização do conhecimento (SOC) usado para classificar conjuntos de dados e serviços do catálogo.	
dct:hasPart	Um item listado no catálogo.	
dcat:dataset	Uma coleção de dados listados no catálogo. Subpropriedade de dct:hasPart.	
dcat:service	Um <i>site</i> ou ponto de extremidade (<i>endpoint</i>) listado no catálogo. Subpropriedade de dct:hasPart.	
dcat:catalog	Um catálogo cujo conteúdo é de interesse no contexto deste catálogo. Subpropriedade de dct:hasPart.	
dcat:record	Um registro que descreve o cadastro de um único conjunto de dados ou serviço de dados que faz parte do catálogo.	
dct:accessRights	Informações sobre quem pode acessar o recurso ou uma indicação de seu <i>status</i> de segurança.	dcat:Resource

Propriedade	Descrição	Classe
dct:conformsTo	Um padrão estabelecido com o qual o recurso descrito está em conformidade.	
dcat:contactPoint	Informações de contato relevantes para o recurso catalogado. Recomenda-se o uso do <i>vCard Ontology</i> .	
dct:creator	A entidade responsável pela produção do recurso.	
dct:description	Um relato de texto livre do item.	
dct:title	Um nome dado ao item.	
dct:issued	Data da emissão formal (por exemplo, publicação) do item.	
dct:modified	Data mais recente em que o item foi alterado, atualizado ou modificado.	
dct:language	Um idioma do item. Refere-se ao idioma natural usado para metadados textuais (ou seja, títulos, descrições etc.) de um recurso catalogado (ou seja, conjunto de dados ou serviço) ou aos valores textuais de uma distribuição de conjunto de dados.	
dct:Publisher	A entidade responsável por disponibilizar o item.	
dct:identifier	Um identificador único do item.	
dcat:theme	Uma categoria principal do recurso. Um recurso pode ter vários temas. Subpropriedade de dct:subject.	
dct:type	A natureza ou gênero do recurso. Subpropriedade de dc:type.	
dct:relation	Um recurso com um relacionamento não especificado com o item catalogado.	
dcat:qualifiedRelation	Ligação para uma descrição de um relacionamento com outro recurso. Subpropriedade de prov:qualifiedInfluence.	
dcat:keyword	Uma palavra-chave ou etiqueta que descreve o recurso.	
dcat:landingPage	Uma página da <i>Web</i> que pode ser navegada em um navegador da <i>Web</i> para obter acesso ao catálogo, um conjunto de dados, suas distribuições e/ou informações adicionais. Subpropriedade de foaf:page.	
prov:qualifiedAttribution	Ligação para um agente com alguma forma de responsabilidade pelo recurso. Subpropriedade de prov:qualifiedInfluence.	

Propriedade	Descrição	Classe	
dct:license	Um documento legal sob o qual o recurso é disponibilizado.		
dct:rights	Uma declaração que diz respeito a todos os direitos não abordados com dct:license ou dct:accessRights, como declarações de direitos autorais.		
odrl:hasPolicy	Uma política em conformidade com ODRL ¹³ que expressa os direitos associados ao recurso.		
dct:isReferencedBy	Um recurso relacionado, como uma publicação, que faz referência, cita ou aponta para o recurso catalogado.		
dct:title	Um nome dado ao registro.	dcat:CatalogRecord	
dct:description	Um relato de texto livre do registro.		
dct:issued	A data de listagem (ou seja, gravação formal) do conjunto de dados ou serviço correspondente no catálogo.		
dct:modified	Data mais recente em que a entrada do catálogo foi alterada, atualizada ou modificada.		
foaf:primaryTopic	O recurso (dcat:Resource) (conjunto de dados ou serviço) descrito no registro.		
dct:conformsTo	Um padrão estabelecido com o qual o recurso descrito está em conformidade.	dcat:Dataset	
dcat:distribution	Uma distribuição disponível do conjunto de dados. Subpropriedade de dct:relation.		
dct:accrualPeriodicity	A frequência com que o conjunto de dados é publicado.		
dct:spatial	A área geográfica coberta pelo conjunto de dados.		
dcat:spatialResolutionInMeters	Separação espacial mínima resolvível em um conjunto de dados, medido em metros.		
dct:temporal	O período temporal que o conjunto de dados cobre.		
dcat:temporalResolution	Período de tempo mínimo resolvível no conjunto de dados.		
prov:wasGeneratedBy	Uma atividade que gerou, ou fornece o contexto de negócios para, a criação do conjunto de dados.		
dct:title	Um nome dado à distribuição.		dcat:Distribution

¹³ “O *Open Digital Rights Language* (ODRL) é uma linguagem de expressão de política que fornece um modelo de informações, vocabulário e mecanismos de codificação flexíveis e interoperáveis para representar declarações sobre o uso de conteúdo e serviços.” (IANNELLA *et al.*, 2018, n.p., tradução nossa).

Propriedade	Descrição	Classe
dct:description	Um relato de texto livre da distribuição.	
dct:issued	Data da emissão formal (por exemplo, publicação) da distribuição.	
dct:modified	Data mais recente em que a distribuição foi alterada, atualizada ou modificada.	
dct:license	Um documento legal sob o qual a distribuição é disponibilizada.	
dct:accessRights	Uma declaração de direitos que diz respeito a como a distribuição é acessada.	
dct:rights	Informações sobre direitos mantidos na distribuição e sobre ela.	
odrl:hasPolicy	Uma política em conformidade com ODRL que expressa os direitos associados à distribuição.	
dcat:accessURL	Uma URL do recurso que dá acesso a uma distribuição do conjunto de dados. Por exemplo, página de destino, <i>feed</i> , ponto de extremidade SPARQL (SPARQL <i>endpoint</i>).	
dcat:accessService	Um serviço de dados que dá acesso à distribuição do conjunto de dados.	
dcat:downloadURL	A URL do arquivo para <i>download</i> em um determinado formato. Por exemplo, arquivo CSV ou arquivo RDF. O formato é indicado pelo formato (dct:format) da distribuição e/ou tipo de mídia (dcat:mediaType).	
dcat:byteSize	O tamanho de uma distribuição em <i>bytes</i> .	
dcat:spatialResolutionInMeters	A separação espacial mínima resolvível em uma distribuição de conjunto de dados, medida em metros.	
dcat:temporalResolution	Período de tempo mínimo resolvível na distribuição do conjunto de dados.	
dct:conformsTo	Um padrão estabelecido com o qual a distribuição está em conformidade.	
dcat:mediaType	O tipo de mídia da distribuição, conforme definido pela IANA ¹⁴ . Subpropriedade de dct:format.	
dct:format	O formato do arquivo da	

¹⁴ Disponível em: <https://www.iana.org/assignments/media-types/media-types.xhtml>.

Propriedade	Descrição	Classe
	distribuição.	
dcat:compressFormat	O formato de compactação da distribuição no qual os dados estão contidos em uma forma compactada, por exemplo para reduzir o tamanho do arquivo para <i>download</i> .	
dcat:packageFormat	O formato do pacote da distribuição no qual um ou mais arquivos de dados são agrupados, por exemplo para permitir que um conjunto de arquivos relacionados seja baixado juntos.	
dcat:endpointURL	O local raiz ou o ponto de extremidade principal do serviço (um IRI resolvível pela <i>Web</i>).	
dcat:endpointDescription	Uma descrição dos serviços disponíveis pelos pontos de extremidade, incluindo suas operações, parâmetros etc.	dcat:Service
dcat:servesDataset	Uma coleção de dados que este serviço de dados pode distribuir.	
-	-	skos:ConceptScheme
-	-	skos:Concept
-	-	foaf:Person
-	-	foaf:Organization
dct:relation	O recurso relacionado ao recurso de origem.	
dcat:hadRole	A função de uma entidade ou agente em relação a outra entidade ou recurso.	dcat:Relationship
-	-	dcat:Role
dcat:startDate	O início do período.	
dcat:endDate	O fim do período.	
time:hasBeginning	Início de um período ou intervalo de tempo.	dct:PeriodOfTime
time:hasEnd	Fim de um período ou intervalo de tempo.	
locn:geometry	Associa qualquer recurso à geometria correspondente.	
dcat:bbox	A caixa delimitadora geográfica de um recurso.	
dcat:centroid	O centro geográfico (centróide) de um recurso.	dct:Location

Fonte: adaptação e tradução de Albertoni *et al.* (2020).

A partir dos Quadros 15 e 16 é possível constatar a heterogeneidade dos vocabulários que são utilizados para formar o DCAT, sendo muito utilizados elementos do DC *Terms*. Além disso, algumas propriedades se repetem em

diferentes classes, atendendo a cada uma delas em seu contexto, uma vez que o DCAT proporciona a descrição tanto a nível do catálogo quanto do conjunto de dados em si (TEMESIO, 2015). Por exemplo, a propriedade 'dct:title' se repete nas classes 'dcat:Resource', 'dcat:CatalogRecord' e 'dcat:Distribution', sendo que no primeiro é utilizado para indicar nome do item, no segundo, nome do registro, e no terceiro, nome da distribuição.

Adicionalmente, o DCAT é especificado de modo que possa ser aplicado em diferentes domínios do conhecimento a partir de perfis de aplicação, sendo que “Em alguns casos, um perfil estende um dos perfis do DCAT, adicionando classes e propriedades para os campos de metadados não abordados no perfil do DCAT de referência.” (ALBERTONI *et al.*, 2020, n.p., tradução nossa). Segundo Albertoni *et al.* (2020, n.p., tradução nossa), algumas das restrições que um perfil pode apresentar são:

- Restrições de cardinalidade, incluindo um conjunto mínimo de campos de metadados necessários;
- Subclasses e subpropriedades das classes e propriedades padrões do DCAT;
- Classes e propriedades para campos de metadados adicionais não cobertos na especificação de vocabulário DCAT;
- Vocabulários controlados ou conjuntos de URI como valores aceitáveis para propriedades;
- Requisitos para mecanismos de acesso específicos (sintaxes e protocolos RDF) à descrição RDF do catálogo.

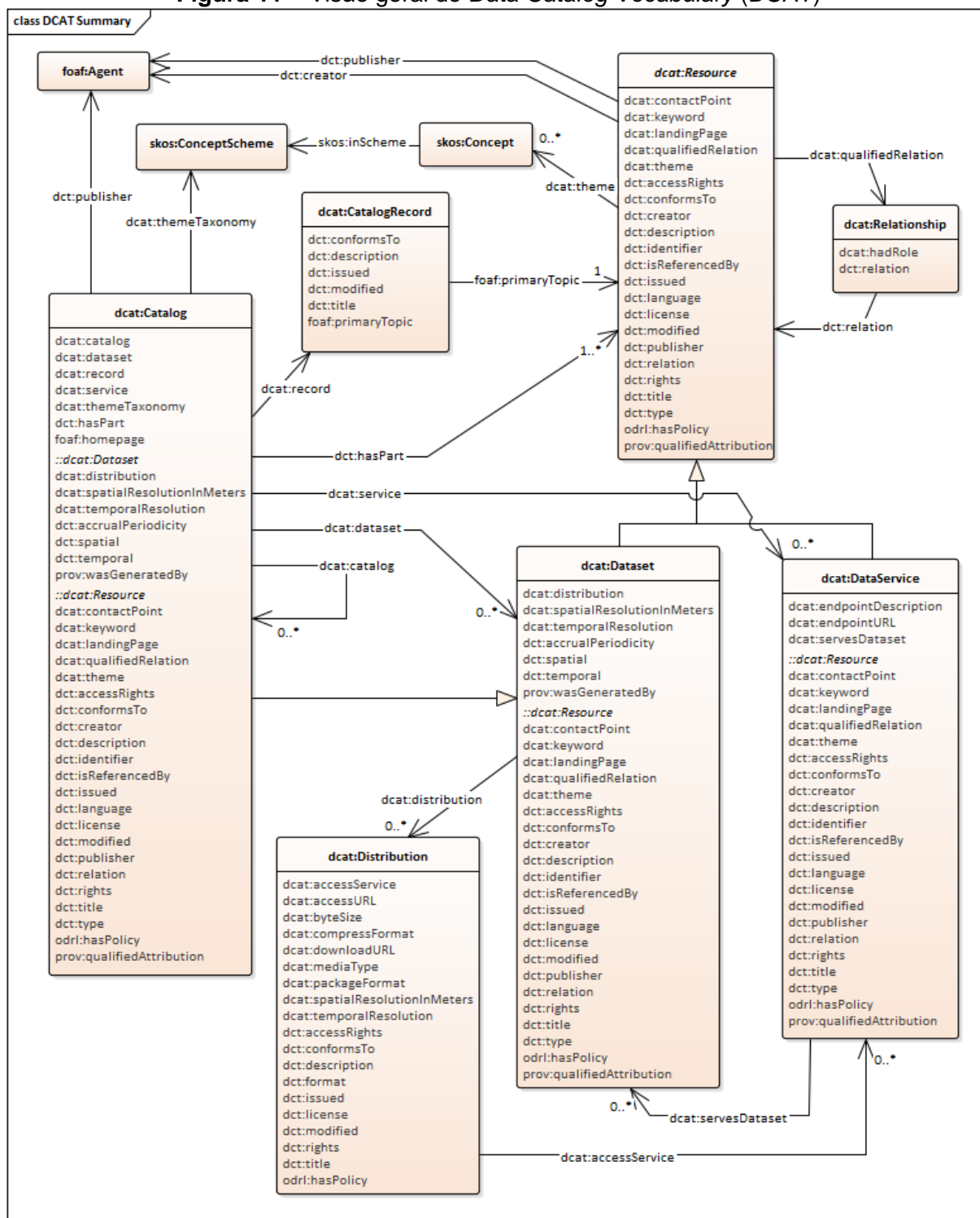
Alguns dos perfis de aplicação do DCAT são:

- **DCAT-AP**: perfil de aplicação do DCAT para portais de dados na Europa;
- **GeoDCAT-AP**: perfil geoespacial de DCAT-AP;
- **StatDCAT-AP**: perfil estatístico de DCAT-AP;
- **DCAT-AP_IT**: perfil italiano de DCAT-AP;
- **GeoDCAT-AP_IT**: perfil italiano de GeoDCAT-AP;
- **DCAT-AP-NO**: perfil norueguês de DCAT-AP;
- **DCAT-AP.de**: perfil alemão de DCAT-AP;

- **DCAT-BE:** perfil belga de DCAT-AP;
- **DCAT-AP-SE:** perfil sueco de DCAT-AP. (ALBERTONI *et al.*, 2020, n.p., tradução e grifo nosso).

A seguir, na Figura 11, apresenta-se uma visão geral do DCAT:

Figura 11 – Visão geral do Data Catalog Vocabulary (DCAT)



Fonte: Albertoni *et al.* (2020).

Na Figura 11 é possível observar como as classes e propriedades do vocabulário se relacionam entre si. Albertoni *et al.* (2020) apresentam alguns exemplos de uso do DCAT. No Quadro 17 abaixo pode ser verificado a descrição (dct:description), o identificador (dct:identifier), o criador (dct:creator) e as relações (dct:relation) de um conjunto de dados.

Quadro 17 – Exemplo de uso do *Data Catalog Vocabulary* (DCAT) em catálogo de dados

```
:d33937
  dct:description "A set of RDF graphs representing the International
[Chrono]stratigraphic Chart, ..."@en ;
  dct:identifier "https://doi.org/10.25919/5b4d2b83cbf2d"^^xsd:anyURI ;
  dct:creator <https://orcid.org/0000-0002-3884-3420>;
  dct:relation <https://vocabs.ands.org.au/viewById/196> ;
  dct:relation :ChronostratChart2017-02.pdf ;
  dct:relation :ChronostratChart2017-02.jpg ;
  dct:relation :timescale.zip ;
  dct:relation :isc2017.jsonld ;
  dct:relation :isc2017.nt ;
  dct:relation :isc2017.rdf ;
  dct:relation :isc2017.ttl ;
.
```

Fonte: Albertoni *et al.* (2020).

Indica-se que a propriedade ‘dct:relation’ deve ser utilizada quando “[...] a natureza dos relacionamentos entre um conjunto de dados e recursos de componentes em um catálogo, repositório ou outro local não é conhecida [...]” (ALBERTONI *et al.*, 2020, n.p., tradução nossa). Já quando se sabe claramente que os recursos relacionados são, de fato, uma representação do conjunto de dados, utiliza-se a propriedade ‘dcat:distribution’.

No Quadro 18 a seguir encontra-se exemplo de indicação de proveniência de um conjunto de dados, a partir de utilização do PROV-O, apontando o que o gerou (prov:wasGeneratedBy) e a descrição dessa atividade (prov:Activity).

Quadro 18 – Exemplo de uso do *Data Catalog Vocabulary* (DCAT), com indicação de proveniência

```
dap:atnf-P366-2003SEPT
  rdf:type dcat:Dataset ;
  dct:bibliographicCitation "Burgay, M; McLaughlin, M; Kramer, M; Lyne, A;
Joshi, B; Pearce, G; D'Amico, N; Possenti, A; Manchester, R; Camilo, F
(2017): Parkes observations for project P366 semester 2003SEPT. v1. CSIRO.
Data Collection. https://doi.org/10.4225/08/598dc08d07bb7" ;
  dct:title "Parkes observations for project P366 semester 2003SEPT"@en ;
  dcat:landingPage <https://data.csiro.au/dap/landingpage?pid=csiro:P366-
2003SEPT> ;
  prov:wasGeneratedBy dap:P366 ;
```

```

dap:P366
  rdf:type prov:Activity ;
  dct:type <http://dbpedia.org/resource/Observation> ;
  prov:startedAtTime "2000-11-01"^^xsd:date ;
  prov:used dap:Parkes-radio-telescope ;
  prov:wasInformedBy dap:ATNF ;
  rdfs:label "P366 - Parkes multibeam high-latitude pulsar survey"@en ;
  rdfs:seeAlso <https://doi.org/10.1111/j.1365-2966.2006.10100.x> ;

```

Fonte: W3C *Dataset Exchange Working Group* (c2020).

Os conjuntos de dados podem ser ligados a publicações a partir da propriedade 'dct:isReferencedBy', conforme o Quadro 19, em que o DOI da publicação é indicado.

Quadro 19 – Exemplo de uso do *Data Catalog Vocabulary* (DCAT), com ligação entre conjunto de dados e publicação

```

:globtherm
  dct:title "Data from: GlobTherm, a global database on thermal tolerances
for aquatic and terrestrial organisms"@en ;
  dct:description "How climate affects species distributions is a
longstanding question receiving renewed interest owing to the need to
predict the impacts of global warming on biodiversity. Is climate change
forcing species to live near their critical thermal limits? Are these
limits likely to change through natural selection? These and other
important questions can be addressed with models relating geographical
distributions of species with climate data, but inferences made with these
models are highly contingent on non-climatic factors such as biotic
interactions. Improved understanding of climate change effects on species
will require extensive analysis of thermal physiological traits, but such
data are scarce and scattered. To overcome current limitations, we created
the GlobTherm database. The database contains experimentally derived
species' thermal tolerance data currently comprising over 2,000 species of
terrestrial, freshwater, intertidal and marine multicellular algae, pl
ants, fungi, and animals. The GlobTherm database will be maintained and
curated by iDiv with the aim of expanding it, and enable further
investigations on the effects of climate on the distribution of life on
Earth."@en ;
  dct:identifier "https://doi.org/10.5061/dryad.1cv08"^^xsd:anyURI ;
  dct:creator <https://orcid.org/0000-0002-7883-3577> ;
  dct:relation <https://doi.org/10.5061/dryad.1cv08/6> ;
  dct:relation <https://doi.org/10.5061/dryad.1cv08/7> ;
  dct:isReferencedBy <https://doi.org/10.1038/sdata.2018.22>.

```

Fonte: W3C *Dataset Exchange Working Group* (c2020).

No Quadro 20 abaixo encontra-se presente exemplo de descrição de serviço de dados (dcat:DataService), no caso, de um catálogo de dados da *European Environment Agency* (EEA).

Quadro 20 – Exemplo de uso do *Data Catalog Vocabulary* (DCAT) para descrição de serviços de dados

```

a:EEA-CSW-Endpoint
  rdf:type dcat:DataService ;
  dct:type <http://inspire.ec.europa.eu/metadata-
codelist/SpatialDataServiceCategory/infoCatalogueService> ;
  dct:accessRights
<http://publications.europa.eu/resource/authority/access-right/PUBLIC> ;
  dct:conformsTo <http://www.opengis.net/def/serviceType/ogc/csw> ;
  dct:description "The EEA public catalogue of spatial datasets references
the spatial datasets used by the European Environment Agency as well as
the spatial datasets produced by or for the EEA. In the latter case,
when datasets are publicly available, a link to the location from where
they can be downloaded is included in the dataset's metadata. The
catalogue has been initially populated with the most important spatial
datasets already available on the data&maps section of the EEA website
and is currently updated with any newly published spatial
dataset."@en ;
  dct:identifier "eea-sdi-public-catalogue" ;
  dct:issued "2012-01-01"^^xsd:date ;
  dct:license <https://creativecommons.org/licenses/by/2.5/dk/> ;
  dct:spatial [
    rdf:type dct:Location ;
    dcat:bbox "POLYGON((-180 90,180 90,180 -90,-180 -90,-180
90))"^^gsp:wktLiteral ;
  ] ;
  dct:title "European Environment Agency's public catalogue of spatial
datasets."@en ;
  dct:type <http://inspire.ec.europa.eu/metadata-
codelist/ResourceType/service> ;
  dct:type <http://inspire.ec.europa.eu/metadata-
codelist/SpatialDataServiceType/discovery> ;
  dcat:contactPoint a:EEA ;
  dcat:endpointDescription
<https://sdi.eea.europa.eu/catalogue/srv/eng/csw?service=CSW&request=GetCap
abilities> ;
  dcat:endpointURL <http://sdi.eea.europa.eu/catalogue/srv/eng/csw> ;
.

```

Fonte: W3C *Dataset Exchange Working Group* (c2020).

Por último, no Quadro 21 está presente um exemplo de descrição de distribuição comprimida em pacote, em GZIP.

Quadro 21 – Exemplo de uso do *Data Catalog Vocabulary* (DCAT) para descrição de distribuições comprimidas e em pacote

```

@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .

<https://data.gov.cz/zdroj/datová-sada/247025684/22> a dcat:Distribution ;
  dcat:accessURL <https://mvcr1.opendata.cz/czechpoint/2007.csv.gz> ;
  dcat:downloadURL <https://mvcr1.opendata.cz/czechpoint/2007.csv.gz> ;
  dct:license <https://data.gov.cz/podminky-užití/volný-přístup/> ;
  dct:conformsTo <https://mvcr1.opendata.cz/czechpoint/2007.json> ;
  dct:format <http://publications.europa.eu/resource/authority/file-
type/CSV> ;
  dcat:mediaType <http://www.iana.org/assignments/media-types/text/csv> ;
  dcat:compressFormat <http://www.iana.org/assignments/media-

```

```
types/application/gzip>
```

Fonte: W3C *Dataset Exchange Working Group* (c2020).

Observa-se que o DCAT se apresenta explicitamente como um vocabulário direcionado à publicação de dados para catálogos, reutilizando-se de elementos advindos de outros vocabulários, promovendo a sustentabilidade. Dessa forma, o DCAT pode ser uma opção de vocabulário para a padronização da descrição de conjuntos de dados de pesquisa, bem como no contexto *Linked Data*.

Aponta-se, também, que o DCAT considera a promoção da ligação de dados em suas recomendações, como ao recomendar que seja evitado o uso de *blank nodes* em descrições em RDF, visto que, apesar de proporcionar flexibilidade, “Um blank node não pode ser o destino de um link e não pode ser anotado com novas informações de novas fontes.” (ALBERTONI *et al.*, 2020, n.p., tradução nossa).

Desse modo, partindo para uma análise dos vocabulários no contexto dos dados de pesquisa, em 2013 Qin e Li buscaram compreender quão portátil são os padrões de metadados para os dados de pesquisa, ou seja, qual “[...] a habilidade de elementos semânticos em um padrão de metadados para ser reutilizado em diferentes contextos por meio de aplicações interoperáveis.” (QIN; LI, 2013, p. 27, tradução nossa).

O estudo analisou 16 padrões de metadados: *Access to Biological Collection Data* (ABCD), *Astronomy Visualization Metadata Standard* (AVM), *ClinicalTrials.gov Protocol Data Element Definitions* (Clinical), *Content Standard for Digital Geospatial Metadata* (CSDGM), *Content Standard for Digital Geospatial Metadata, Part 1: Biological Data Profile* (CSDGM-BD), *Darwin Core* (Darwin), *Dublin Core Metadata Element Set* (DCMES), *Ecological Metadata Language* (EML), *GenBank Flat File Format* (GenBank), *Genome Metadata* (Genome), *International Virtual Observatory Alliance* (IVOA), *ISO/TS 19115:2003 Geographic information - Metadata* (ISO 19115), *Metadata Profile for Shoreline Data* (CSDGM-SL), *NetCDF Climate and Forecast Metadata Convention* (CF), *NISO Metadata for Images in XML* (NISO-Image) e *WHO Trial Registration Data Set* (TRDS) (QIN; LI, 2013).

Apesar do DCMES ser um vocabulário de uso mais genérico, os autores verificaram que a maioria dos elementos e categorias levantados por eles condizem com os do DCMES, bem como *DataCite* (QIN; LI, 2013).

De modo semelhante, Farnel e Shiri (2014) analisam os repositórios *DataCite*, *Dataverse Network*, *Dryad* e *FigShare* a fim de identificar os metadados utilizados por eles. Os autores identificam que diversos padrões são utilizados, mas que o *Dublin Core* é utilizado ou suportado pela maior parte deles. A partir disso, observa-se a grande adoção de termos *Dublin Core* para a descrição de dados de pesquisa (ALAM, 2014), e aponta que o vocabulário proporciona, efetivamente, o reuso e a interoperabilidade. Encontra-se, desse modo, a motivação do DCAT de adotar um grande número de termos advindo do *DC Terms*.

Retoma-se que o DCAT faz grande uso de termos de outros vocabulários na uniformização dos conjuntos de dados, de forma a reaproveitar termos já existentes em sua construção, como os termos do *DC Terms*, *FOAF Vocabulary*, *vCard Ontology* e *PROV-O*, além da possibilidade de adoção do *VOID* (ALBERTONI *et al.*, 2020; BELTRAN *et al.*, 2018), especificados nas subseções a seguir.

5.1 Dublin Core Metadata Initiative Metadata Terms (DC Terms)

Com criação em 1995, o *Dublin Core* (ou *Dublin Core Metadata Element Set - DCMES*) foi desenvolvido com o objetivo de fornecer metadados para a padronização da descrição de recursos informacionais eletrônicos, visando a sua recuperação (DCMI USAGE BOARD, 2020; WEIBEL *et al.*, 1998). Conforme menciona Alves (2010, p. 64), “O padrão de metadados Dublin Core - DC - se constitui em um simples e eficaz esquema de metadados utilizado para descrever uma ampla variedade de recursos em diversos ambientes informacionais da Web.”.

São quinze os metadados do DCMES: título (*title*), autor (*creator*), assunto (*subject*), descrição (*description*), editor (*publisher*), colaborador (*contributor*), data (*date*), tipo (*type*), formato (*format*), identificador (*identifier*), fonte (*source*), língua (*language*), relação (*relation*), cobertura (*coverage*) e direitos (*rights*) (DCMI USAGE BOARD, 2020; WEIBEL *et al.*, 1998).

O esquema *Dublin Core* é mantido pela organização aberta *Dublin Core Metadata Initiative* (DCMI), que objetiva apoiar o compartilhamento de metadados e a inovação, por meio de:

- Gerenciamento de curadoria a longo prazo e desenvolvimento de especificações DCMI e namespaces de termos de metadados;
- Gerenciando a discussão contínua dos temas de trabalho atuais da DCMI;
- Configuração e gestão de eventos internacionais e regionais;

- Curadoria e disponibilidade aberta de recursos para reuniões, incluindo procedimentos, relatórios de projetos e atas de reuniões;
- Criação e entrega de recursos de treinamento nas melhores práticas de metadados, incluindo tutoriais, webinars e workshops; e
- Coordenação da comunidade global de voluntários da DCMI. (DUBLIN CORE METADATA INITIATIVE, c2020a, n.p., tradução nossa).

Com o desenvolvimento da *Web* a noção de metadados *Dublin Core* se expandiu, sendo hoje “[...] metadados projetados para interoperabilidade com base na *Web Semântica* ou nos princípios *Linked Data*.” (DUBLIN CORE METADATA INITIATIVE, c2020b, n.p., tradução nossa).

O *Dublin Core* pode ser utilizado em conjunto com propriedades, classes, *datatypes* e esquemas de codificação de vocabulário adicionais, formando o chamado *DC Terms*, sendo que “Esses termos devem ser usados em combinação com termos de metadados de outros vocabulários compatíveis no contexto dos perfis de aplicação.” (DCMI USAGE BOARD, 2020, n.p., tradução nossa).

Um perfil de aplicação consiste em um “[...] conjunto de elementos de metadados, políticas e diretrizes definidos para uma aplicação ou implementação em particular.”, sendo que pode ser formado por elementos de diferentes esquemas (WOODLEY; CLEMENT; WINN, 2005, n.p., tradução nossa).

De acordo com Heery e Patel (2000, n.p., tradução nossa),

Os perfis de aplicação consistem em elementos de dados extraídos de um ou mais esquemas de namespace combinados por implementadores e otimizados para uma aplicação local específica. Os perfis de aplicação são úteis, pois permitem que o implementador declare como eles estão usando esquemas padrão.

Portanto, os perfis de aplicação são desenvolvidos com o propósito de estender os elementos de um esquema pré-existente, de modo a servir melhor em contextos mais específicos, visando o aprimoramento da representação da informação (CERRAO, 2019).

Nesse contexto, o DCMI desenvolveu o *Dublin Core Application Profile* (DCAP), modelo que, de acordo com Coyle e Baker (2009, n.p., tradução nossa):

- descreve o que uma comunidade deseja realizar com sua aplicação (Requisitos Funcionais);
- caracteriza os tipos de coisas descritas pelos metadados e seus relacionamentos (Modelo de Domínio);

- enumera os termos de metadados a serem usados e as regras para seu uso (Perfil de Conjunto de Descrição e Diretrizes de Uso); e
- define a sintaxe da máquina que será usada para codificar os dados (Diretrizes de Sintaxe e Formatos de Dados).

O DCAP segue o DCMI *Abstract Model* (DCAM) (COYLE; BAKER, 2009), modelo abstrato da DCMI, desenvolvido sobre o RDF com o objetivo de prover um modelo informacional que especifica “[...] a natureza dos componentes usados e descreve como esses componentes são combinados para criar estruturas de informação.” (POWELL *et al.*, 2007, n.p., tradução nossa).

Observa-se que maiores discussões sobre perfis de aplicação estão presentes em autores como Malta e Baptista (2013; 2014), em que são explorados, respectivamente, a proposta de um método para o desenvolvimento de perfis de aplicação *Dublin Core* (Me4DCAP V0.2) e um mapeamento da utilização dos perfis de aplicação no geral, com base em revisão de literatura. Adicionalmente, o DCMI DCAP *Interest Group* tem trabalhado no projeto de desenvolvimento de um vocabulário central para perfis (*Core Vocabulary for Profiles*), com a intenção de facilitar a criação de perfis de aplicação (DUBLIN CORE METADATA INITIATIVE, 2020).

Quanto aos termos do *Dublin Core*, o DCMI *Metadata Terms*, também chamado de DC *Terms*, reúne a totalidade desses termos de metadados mantidos pela DCMI, incluindo propriedades, esquemas de codificação de vocabulário, esquemas de codificação de sintaxe e classes, seccionados de acordo com o Quadro 22 (DCMI USAGE BOARD, 2020).

Quadro 22 – Tipos de termos de metadados mantidos pela DCMI

Tipo	Termos
Propriedades no <i>namespace</i> <u>/terms/</u>	abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation, conformsTo, contributor, coverage, created, creator, date, dateAccepted, dateCopyrighted, dateSubmitted, description, educationLevel, extent, format, hasFormat, hasPart, hasVersion, identifier, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, issued, isVersionOf, language, license, mediator, medium, modified, provenance, publisher, references, relation, replaces, requires, rights, rightsHolder, source,

	spatial, subject, tableOfContents, temporal, title, type, valid
Propriedades no <i>namespace</i> <u>/elements/1.1/</u>	contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type
Esquemas de codificação de vocabulário (<i>Vocabulary Encoding Schemes</i>)	DCMIType, DDC, IMT, LCC, LCSH, MESH, NLM, TGN, UDC
Esquemas de codificação de sintaxe (<i>Syntax Encoding Schemes</i>)	Box, ISO3166, ISO639-2, ISO639-3, Period, Point, RFC1766, RFC3066, RFC4646, RFC5646, URI, W3CDTF
Classes	Agent, AgentClass, BibliographicResource, FileFormat, Frequency, Jurisdiction, LicenseDocument, LinguisticSystem, Location, LocationPeriodOrJurisdiction, MediaType, MediaTypeOrExtent, MethodOfAccrual, MethodOfInstruction, PeriodOfTime, PhysicalMedium, PhysicalResource, Policy, ProvenanceStatement, RightsStatement, SizeOrDuration, Standard
Vocabulário do tipo DCMI (DCMI <i>Type Vocabulary</i>)	Collection, Dataset, Event, Image, InteractiveResource, MovingImage, PhysicalObject, Service, Software, Sound, StillImage, Text
Termos para descrição de vocabulário	domainIncludes, memberOf, rangelIncludes, VocabularyEncodingScheme

Fonte: tradução de DCMI *Usage Board* (2020).

Desse modo, é possível estender e especificar a descrição de recursos informacionais ao utilizar os diversos termos do DC *Terms* listados no Quadro 22 acima. No Quadro 23 a seguir apresenta-se exemplo de uso do DC *Terms*:

Quadro 23 – Exemplo de uso do DCMI *Metadata Terms*

<pre><?xml version="1.0"?> <metadata xmlns="http://example.org/myapp/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://example.org/myapp/ http://example.org/myapp/schema.xsd" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/"> <dc:title> UKOLN </dc:title> <dcterms:alternative> UK Office for Library and Information Networking </dcterms:alternative> <dc:subject> national centre, network information support, library community, awareness, research, information services,public library networking, bibliographic management, distributed library systems, metadata, resource discovery,</pre>

```

    conferences,lectures, workshops
  </dc:subject>
  <dc:subject xsi:type="dcterms:DDC">
    062
  </dc:subject>
  <dc:subject xsi:type="dcterms:UDC">
    061(410)
  </dc:subject>
  <dc:description>
    UKOLN is a national focus of expertise in digital information
    management. It provides policy, research and awareness services
    to the UK library, information and cultural heritage communities.
    UKOLN is based at the University of Bath.
  </dc:description>
  <dc:description xml:lang="fr">
    UKOLN est un centre national d'expertise dans la gestion de
    l'information
    digitale.
  </dc:description>
  <dc:publisher>
    UKOLN, University of Bath
  </dc:publisher>
  <dcterms:isPartOf xsi:type="dcterms:URI">
    http://www.bath.ac.uk/
  </dcterms:isPartOf>
  <dc:identifi er xsi:type="dcterms:URI">
    http://www.ukoln.ac.uk/
  </dc:identifi er>
  <dcterms:modified xsi:type="dcterms:W3CDTF">
    2001-07-18
  </dcterms:modified>
  <dc:format xsi:type="dcterms:IMT">
    text/html
  </dc:format>
  <dcterms:extent>
    14 Kbytes
  </dcterms:extent>
</metadata>

```

Fonte: Powell e Johnston (2003).

O exemplo acima consiste em um registro que se utiliza do vocabulário DC *Terms*, estando escrito em XML e utilizando-se do XML *Schema*. Nele pode ser observado a utilização de termos dos tipos: ‘Propriedades no *namespace /terms/*’ (dcterms:alternative, dcterms:isPartOf, dcterms:modified, dcterms:extent); ‘Propriedades no *namespace /elements/1.1/*’ (dc:title, dc:subject, dc:description, dc:publisher, dc:identifi er, dc:format); ‘Esquemas de codificação de vocabulário’ (dcterms:DDC, dcterms:UDC, dcterms:IMT); e ‘Esquemas de codificação de sintaxe’ (dcterms:URI, dcterms:W3CDTF). Com a adoção de diferentes tipos de termos é possível apresentar o registro com um detalhamento maior, tornando a sua descrição mais completa.

5.2 *Friend of a Friend Vocabulary (FOAF Vocabulary)*

O *Friend of a Friend* (FOAF) é uma linguagem desenvolvida por Dan Brickley e Libby Miller que define os termos relacionados às pessoas que podem ser utilizados em dados estruturados (BRICKLEY; MILLER, 2014). O projeto objetiva-se a interligar pessoas no âmbito da *Web*, fornecendo “[...] uma abordagem em que diferentes sites podem contar diferentes partes de um cenário maior, e através do qual os usuários podem manter algum controle sobre suas informações em um formato não proprietário.” (BRICKLEY; MILLER, 2014, n.p., tradução nossa).

O FOAF utiliza-se de ferramentas semânticas, como o RDF e RDFa, e vocabulários, como esquemas e ontologias, para garantir que os documentos FOAF formados – descrições FOAF publicadas na *Web* – sejam compreensíveis por máquinas. Os documentos FOAF devem, ainda, ser documentos RDF (BRICKLEY; MILLER, 2014).

O conjunto de termos utilizados na ligação de pessoas é denominado como *FOAF Vocabulary*, vocabulário escrito em RDF/OWL de termos básicos utilizáveis nas páginas *Web*, identificado pelo *namespace* URI fixo ‘http://xmlns.com/foaf/0.1/’. O vocabulário é formado por classes, compreendidas como “Os tipos de coisas sobre as quais falamos no FOAF [...]”, e propriedades, a “[...] variedade de coisas e ligações [...]” (BRICKLEY; MILLER, 2014, n.p., tradução nossa).

Os termos que compõem o vocabulário são categorizados de acordo com sua estabilidade, em ‘instável’, ‘em teste’ e ‘estável’, que indicam o desenvolvimento do FOAF, enquanto os termos mais antigos são classificados como ‘arcaicos’, sendo que estes não são utilizados, mas “[...] permitem que as formas antigas se tornem modernas novamente.” (BRICKLEY; MILLER, 2014). Destaca-se que a última atualização realizada no documento de especificação do *FOAF Vocabulary* ocorreu em 2014. No Quadro 24 são apresentadas as classes que formam o *FOAF Vocabulary*, acompanhadas de suas definições e *status* em que se encontram.

Quadro 24 – Classes do *Friend of a Friend Vocabulary* (FOAF Vocabulary)

Classe	Definição	Status
foaf:Agent	Um agente (pessoa, grupo, <i>software</i> ou artefato material).	Estável
foaf:Document	Um documento.	Estável
foaf:Group	Representa uma classe de agentes.	Estável
foaf:Image	Representa uma imagem.	Estável
foaf:LabelProperty	Qualquer propriedade RDF com valores textuais que servem de	Instável

	etiquetas.	
foaf:OnlineAccount	Representa uma conta <i>online</i> associada a um agente.	Em teste
foaf:OnlineChatAccount	É um foaf:OnlineAccount dedicado ao chat/mensagem instantânea.	Instável
foaf:OnlineEcommerceAccount	É um foaf:OnlineAccount dedicado a compra/venda de bens, serviços etc.	Instável
foaf:OnlineGamingAccount	É um foaf:OnlineAccount dedicado a jogos <i>online</i> .	Instável
foaf:Organization	Representa uma organização (companhias, sociedades etc.).	Estável
foaf:Person	Representa uma pessoa.	Estável
foaf:PersonalProfileDocument	Representa documentos que usam RDF para descrever as propriedades da pessoa que os criou.	Em teste
foaf:Project	Representa os projetos (um esforço coletivo de algum tipo).	Em teste

Fonte: adaptação e tradução de Brickley e Miller (2014).

No Quadro 25 encontram-se as propriedades que constituem o FOAF *Vocabulary*.

Quadro 25 – Propriedades do *Friend of a Friend Vocabulary* (FOAF *Vocabulary*)

Propriedade	Definição	Status
foaf:homepage	Relaciona algo a uma página sobre ele.	Estável
foaf:isPrimaryTopicOf	Relaciona algo a um documento que seja principalmente sobre ele.	Estável
foaf:knows	Relaciona uma pessoa (foaf:Person) a outra pessoa (foaf:Person) que ela conheça.	Estável
foaf:made	Relaciona um agente (foaf:Agent) a algo feito por ele.	Estável
foaf:maker	Relaciona algo a um agente (foaf:Agent) que o tenha feito.	Estável
foaf:mbox	Relacionamento entre o dono da caixa de correio e a caixa de correio.	Estável
foaf:member	Relaciona um grupo (foaf:Group) a um agente (foaf:Agent) membro desse grupo.	Estável
foaf:page	Relaciona algo a um documento sobre ele.	Estável
foaf:primaryTopic	Relaciona um documento ao assunto principal do documento.	Estável
foaf:weblog	Relaciona um agente (foaf:Agent) a um <i>weblog</i> (ou <i>blog</i>) desse agente.	Estável
foaf:account	Relaciona um agente (foaf:Agent) a uma conta <i>online</i> (foaf:OnlineAccount) da qual ele é o titular único.	Em teste
foaf:accountName	Indica o nome (identificador) associado a uma conta <i>online</i> (foaf:OnlineAccount).	Em teste

Propriedade	Definição	Status
foaf:accountServiceHomepage	Indica o relacionamento entre uma conta <i>online</i> (foaf:OnlineAccount) à página do provedor de serviços de suporte.	Em teste
foaf:aimChatID	Relaciona um agente (foaf:Agent) a um identificador textual (screenname) designado a ele no sistema AOL Instant Messenger (AIM).	Em teste
foaf:based_near	Relaciona duas coisas espaciais (locais) segundo uma noção humana de proximidade.	Em teste
foaf:currentProject	Relaciona uma pessoa (foaf:Person) a um documento (foaf:Document) indicando um empreendimento colaborativo ou individual.	Em teste
foaf:depiction	Relacionamento entre uma coisa e a imagem (foaf:Image) que a retrata.	Em teste
foaf:depicts	Relacionamento entre uma imagem (foaf:Image) e alguma coisa que a imagem retrata.	Em teste
foaf:familyName	Indica o nome familiar de uma pessoa.	Em teste
foaf:firstName	Indica o primeiro nome de uma pessoa.	Em teste
foaf:focus	Relacionamento entre a conceitualização de uma coisa e a coisa em si.	Em teste
foaf:gender	Relaciona um agente (foaf:Agent), geralmente uma pessoa (foaf:Person), a uma cadeia de caracteres representando seu gênero.	Em teste
foaf:givenName	Indica o nome dado a uma pessoa.	Em teste
foaf:icqChatID	Relaciona um agente (foaf:Agent) a um identificador textual designado a ele no sistema ICQ Chat.	Em teste
foaf:img	Relaciona uma pessoa (foaf:Person) a uma imagem (foaf:Image) que a representa.	Em teste
foaf:interest	Representa o interesse de um agente (foaf:Agent) por meio da indicação de um documento (foaf:Document) cujo tópico caracterize amplamente seu interesse.	Em teste
foaf:jabberID	Relaciona um agente (foaf:Agent) a um identificador textual designado a ele no sistema de mensagens Jabber.	Em teste
foaf:lastName	Indica o sobrenome de uma pessoa.	Em teste
foaf:logo	Indica um logotipo que representa alguma coisa.	Em teste
foaf:mbox_sha1sum	Representação textual do resultado da aplicação de SHA1 ao identificador (URI) de 'mailto:' para uma caixa de correios na <i>Internet</i> .	Em teste
foaf:msnChatID	Relaciona um agente (foaf:Agent) a um identificador textual designado a ele no	Em teste

Propriedade	Definição	Status
	sistema de <i>chat online</i> da <i>Microsoft</i> .	
foaf:myersBriggs	Representa a classificação de personalidade Myers Briggs (MBTI).	Em teste
foaf:name	Indica o nome de alguma coisa.	Em teste
foaf:nick	Relaciona uma pessoa (foaf:Person) a um apelido.	Em teste
foaf:openid	Indica o OpenID de um agente (foaf:Agent).	Em teste
foaf:pastProject	Indica um projeto no qual uma pessoa trabalhou anteriormente.	Em teste
foaf:phone	Indica um telefone.	Em teste
foaf:plan	Indica um comentário .plan (arquivo .plan).	Em teste
foaf:publications	Indica as publicações associadas a uma pessoa (foaf:Person).	Em teste
foaf:schoolHomepage	Relaciona uma pessoa (foaf:Person) a um documento (foaf:Document) que é a página (foaf:homepage) de uma escola que ela frequentou.	Em teste
foaf:skypeID	Relaciona um agente (foaf:Agent) a um nome de conta no Skype.	Em teste
foaf:thumbnail	Relacionamento entre uma imagem (foaf:Image) em tamanho real e uma menor que a representa.	Em teste
foaf:tipjar	Relaciona um agente (foaf:Agent) a um documento (foaf:Document) que descreve alguns mecanismos para pagamento ou recompensa ao agente.	Em teste
foaf:title	Indica o título (Sr., Sra, Srta., Dr., etc.).	Em teste
foaf:topic	Relaciona um documento (foaf:Document) a alguma coisa sobre o assunto do documento.	Em teste
foaf:topic_interest	Relaciona um agente (foaf:Agent) diretamente a alguma coisa sobre a qual ele tenha interesse.	Em teste
foaf:workInfoHomepage	Indica a página de informações de trabalho de uma pessoa (foaf:Person).	Em teste
foaf:workPlaceHomepage	Indica a página de local de trabalho da organização (foaf:Organization) que uma pessoa (foaf:Person) trabalha.	Em teste
foaf:yahooChatID	Relaciona um agente (foaf:Agent) a um identificador textual designado a ele no sistema de <i>chat online</i> do Yahoo.	Em teste
foaf:age	Indica a idade em anos de um agente (foaf:Agent).	Instável
foaf:birthday	Indica a data de nascimento de um agente (foaf:Agent), em cadeia de caracteres no formato 'mm-dd'.	Instável
foaf:membershipClass	Indica a classe de indivíduos que são membros de um grupo (foaf:Group).	Instável
foaf:sha1	Um <i>hash</i> (uma representação) sha1sum, em hexadecimal.	Instável
foaf:status	Indica uma cadeia de caracteres que	Instável

Propriedade	Definição	Status
	expressa sobre o que o usuário está contente para o público geral saber sobre suas atividades atuais.	
foaf:dnaChecksum	Uma 'piada' para lembrar que existem muitas propriedades para identificar as pessoas.	Arcaica
foaf:family_name	Grafia arcaica de foaf:familyName.	Arcaica
foaf:fundedBy	Indica uma organização financiando um projeto ou pessoa.	Arcaica
foaf:geekcode	Representa o 'Geek Code' de uma pessoa (foaf:Person).	Arcaica
foaf:givenname	Grafia arcaica de foaf:givenName.	Arcaica
foaf:holdsAccount	Grafia arcaica de foaf:Account.	Arcaica
foaf:surname	Indica o sobrenome de uma pessoa.	Arcaica
foaf:theme	Indica um tema.	Arcaica

Fonte: adaptação e tradução de Brickley e Miller (2014).

Nota-se que a maior parte das propriedades se encontra em fase de teste, significando que sua utilização ainda não é consistente.

Como exemplo de utilização do vocabulário, Brickley e Miller (2014) apresentam um trecho escrito em FOAF conforme a Figura 12 abaixo.

Figura 12 – Exemplo de uso do FOAF Vocabulary

```
<foaf:Person rdf:about="#danbri" xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:name>Dan Brickley</foaf:name>
  <foaf:homepage rdf:resource="http://danbri.org/" />
  <foaf:openid rdf:resource="http://danbri.org/" />
  <foaf:img rdf:resource="/images/me.jpg" />
</foaf:Person>
```

Fonte: Brickley e Miller (2014).

Basicamente, na Figura 12 consta que: existe uma pessoa (foaf:Person) com um nome (foaf:name) que é 'Dan Brickley'; essa pessoa está em uma homepage (foaf:homepage) e um OpenID (foaf:openid) relacionados a uma coisa denominada como http://danbri.org e em uma imagem (foaf:img) relacionada a uma coisa referenciada pelo URI relativo /images/me.jpg. (BRICKLEY; MILLER, 2014).

Tendo sido criado para uso na *Web*, o FOAF está ligado à *Web Semântica*, com objetivos similares de prover um ambiente de informações ligáveis, sendo um sistema *Linked Data* de dados relacionados principalmente a pessoas. O FOAF busca "[...] integrar informações factuais com informações em documentos orientados para o ser humano [...], bem como informações que ainda estão na

cabeça das pessoas.”, garantindo um escopo de dados amplo, além da riqueza das ligações (BRICKLEY; MILLER, 2014).

5.3 vCard Ontology

O *vCard Ontology* tem como objetivo “[...] promover o uso do vCard para a descrição de pessoas e organizações que utilizam técnicas da web semântica e permitir compatibilidade com implementações tradicionais do vCard.” (IANNELLA; MCKINNEY, 2014, n.p., tradução nossa). Desse modo, busca-se fornecer um vocabulário para que se possa padronizar as representações semânticas em *vCard*.

Conforme Perreault (2011, n.p., tradução nossa) coloca,

Os catálogos de endereços eletrônicos tornaram-se onipresentes. Sua presença crescente em dispositivos portáteis conectados, bem como a diversidade de plataformas que trocam dados de contato exigem um padrão.

Desse modo, o *vCard* é um formato de dados desenvolvido pelo *Internet Engineering Task Force vCard Working Group* pensando-se na representação e intercâmbio de informações relacionadas a pessoas e entidades, permitindo “[...] capturar e trocar informações normalmente armazenadas em agenda ou aplicação de diretório.” (PERREAULT, 2011, n.p., tradução nossa).

Os termos que formam seu vocabulário são classificados em classes, propriedades de objeto (*object properties*), propriedades de dado (*data properties*) e declarações de espaço de nomes (*namespace declarations*). No Quadro 26 a seguir são apresentadas as classes do *vCard Ontology*, acompanhadas, quando existirem, de suas descrições, informação de sua atual continuidade e suas superclasses. Para facilitar a identificação, as classes obsoletas encontram-se em cinza.

Quadro 26 – Classes do vCard Ontology

Classe	Descrição	Superclasse
vcard:Acquaintance	-	Relation Type
vcard:Address	Para especificar os componentes do endereço de entrega para o objeto.	-
vcard:Agent	-	Relation Type
vcard:BBS	Classe obsoleta.	Phone
vcard:Car	Classe obsoleta.	Phone
vcard:Cell	Também chamado de telefone celular.	Phone
vcard:Child	-	Relation Type
vcard:Colleague	-	Relation Type
vcard>Contact	-	Relation Type

Classe	Descrição	Superclasse
vcard:Coresident	-	Relation Type
vcard:Coworker	-	Relation Type
vcard:Crush	-	Relation Type
vcard>Date	-	Relation Type
vcard:Dom	Classe obsoleta.	Type
vcard:Email	Para especificar o endereço de correio eletrônico para comunicação com o objeto que o <i>vCard</i> representa. Use a propriedade do objeto 'hasEmail'.	-
vcard:Emergency	-	Relation Type
vcard:Fax	-	Phone
vcard:Female	-	Gender
vcard:Friend	-	Relation Type
vcard:Gender	Usado para códigos de gênero. O URI do código de gênero deve ser usado como o valor para 'Gender'.	-
vcard:Group	Objeto representando um grupo de pessoas ou entidades. Um objeto de grupo geralmente contém propriedades 'hasMember' para especificar os membros do grupo.	Kind
vcard:Home	Isso implica que a propriedade esteja relacionada à vida pessoal de um indivíduo.	Type
vcard:Individual	Um objeto que representa uma única pessoa ou entidade.	Kind
vcard:Internet	Classe obsoleta.	Type
vcard:Intl	Classe obsoleta.	Type
vcard:ISDN	Classe obsoleta.	Type
vcard:Kin	-	Relation Type
vcard:Kind	A classe pai (<i>parent class</i>) para todos os objetos.	-
vcard:Label	Classe obsoleta.	Type
vcard:Location	Um objeto que representa um local geográfico nomeado.	Kind
vcard:Male	-	Gender
vcard:Me	-	Relation Type
vcard:Met	-	Relation Type
vcard:Modem	Classe obsoleta.	Phone
vcard:Msg	Classe obsoleta.	Phone
vcard:Muse	-	Relation Type
vcard:Name	Para especificar os componentes do nome do objeto.	-
vcard:Neighbor	-	Relation Type
vcard:None	-	Gender
vcard:Organization	Um objeto que representa uma organização. Uma organização é uma entidade única e pode representar uma empresa ou governo, um departamento ou divisão dentro de uma empresa ou governo, um clube, uma associação ou algo semelhante.	Kind
vcard:Other	-	Gender

Classe	Descrição	Superclasse
vcard:Pager	-	Phone
vcard:Parcel	Classe obsoleta.	Type
vcard:Parent	-	Relation Type
vcard:PCS	Classe obsoleta.	Phone
vcard:Phone	Usado para códigos de tipo de telefone. O URI do código do tipo de telefone deve ser usado como valor para o tipo de telefone.	-
vcard:Postal	Classe obsoleta.	Type
vcard:Pref	Classe obsoleta.	Type
vcard:Relation Type	Usado para códigos de tipo de relação. O URI do código do tipo de relação deve ser usado como o valor para o tipo de relação.	-
vcard:Sibling	-	Relation Type
vcard:Spouse	-	Relation Type
vcard:Sweetheart	-	Relation Type
vcard:Tel	Classe obsoleta. Use a propriedade do objeto 'hasTelephone'.	-
vcard:Text	Também chamado de telefone sms (<i>sms telephone</i>).	Phone
vcard:Text phone	-	Phone
vcard:Type	Usado para códigos de tipo. O URI do código de tipo deve ser usado como o valor para 'Type'.	-
vcard:Unknown	-	Gender.
vcard:VCard	A classe vCard é equivalente à nova classe 'Kind', que é o pai dos quatro tipos explícitos de vCards ('Individual', 'Organization', 'Location', 'Group').	-
vcard:Video	-	Phone
vcard:Voice	-	Phone
vcard:Work	Isso implica que a propriedade está relacionada ao local de trabalho de um indivíduo.	Type
vcard:X400	Classe obsoleta.	Type

Fonte: adaptação e tradução de Iannella e McKinney (2014).

É importante observar que as classes apresentam outras relações, como equivalência a termos de outros vocabulários e de alcance (*range*), sendo que na presente pesquisa optou por listar apenas suas superclasses por essas auxiliarem na interpretação e compreensão da abrangência e especificidade do termo.

No Quadro 27 são listadas as propriedades de objeto com suas respectivas descrições, quando houver, bem como informações de continuidade, sendo que as propriedades obsoletas se encontram em cinza.

Quadro 27 – Propriedades de objeto do vCard Ontology

Propriedades	Descrição
vcard:adress	-
vcard:agent	Propriedade obsoleta.
vcard:email	-
vcard:geo	-
vcard:hasAdditionalName	Usado para suportar parâmetros de propriedade para a propriedade de dados de nome adicional.
vcard:hasAdress	Para especificar os componentes do endereço de entrega para o objeto.
vcard:hasCalendarBusy	Para especificar o horário de funcionamento associado ao objeto.
vcard:hasCalendarLink	Para especificar o calendário associado ao objeto.
vcard:hasCalendarRequest	Para especificar o endereço do usuário do calendário para o qual uma solicitação de agendamento será enviada para o objeto.
vcard:hasCategory	Usado para suportar parâmetros de propriedade para a propriedade de dados da categoria.
vcard:hasCountryName	Usado para suportar parâmetros de propriedade da propriedade de dados do nome do país.
vcard:hasEmail	Para especificar o endereço de correio eletrônico para comunicação com o objeto.
vcard:hasFamilyName	Usado para suportar parâmetros de propriedade da propriedade de dados do nome de família.
vcard:hasFormattedName	Usado para suportar parâmetros de propriedade para a propriedade de dados de nome formatado.
vcard:hasGender	Para especificar a identidade de sexo ou gênero do objeto. Os URIs são recomendados para permitir o uso de códigos interoperáveis de sexo e gênero.
vcard:hasGeo	Especificar informações relacionadas ao posicionamento global do objeto. Também pode ser usado como um parâmetro de propriedade.
vcard:hasGivenName	Usado para suportar parâmetros de propriedade para a propriedade de dados de nome determinado.
vcard:hasHonorificPrefix	Usado para suportar parâmetros de propriedade da propriedade de dados do prefixo honorífico.
vcard:hasHonorificSuffix	Usado para suportar parâmetros de propriedade da propriedade de dados do sufixo honorífico.
vcard:hasKey	Para especificar uma chave pública ou certificado de autenticação associado ao objeto.
vcard:hasLanguage	Usado para suportar parâmetros de propriedade para a propriedade de dados do idioma.
vcard:hasLocality	Usado para suportar parâmetros de propriedade para a propriedade de dados da localidade.
vcard:hasLogo	Para especificar uma imagem gráfica de um logotipo associado ao objeto.
vcard:hasMember	Para incluir um membro no grupo que este objeto representa. (Esta propriedade pode ser usada apenas por indivíduos da classe Grupo).
vcard:hasMessaging	Para especificar as mensagens instantâneas e as comunicações do protocolo de presença com o objeto.
vcard:hasName	Para especificar os componentes do nome do objeto.
vcard:hasNickname	Usado para suportar parâmetros de propriedade para a propriedade de dados de apelido.

Propriedades	Descrição
vcard:hasNote	Usado para suportar parâmetros de propriedade da propriedade de dados da nota.
vcard:hasOrganizationName	Usado para suportar parâmetros de propriedade da propriedade de dados do nome da organização.
vcard:hasOrganizationUnitName	Usado para suportar parâmetros de propriedade da propriedade de dados do nome da unidade de organização.
vcard:hasPhoto	Para especificar uma imagem ou informação de fotografia que comente algum aspecto do objeto.
vcard:hasPostalCode	Usado para suportar parâmetros de propriedade da propriedade de dados do código postal.
vcard:hasRegion	Usado para suportar parâmetros de propriedade para a propriedade de dados da região.
vcard:hasRelated	Para especificar um relacionamento entre outra entidade e a entidade representada por este objeto.
vcard:hasRole	Usado para suportar parâmetros de propriedade para a propriedade de dados de função.
vcard:hasSound	Para especificar uma informação de conteúdo de som digital que comente algum aspecto do objeto
vcard:hasSource	Para identificar a fonte de informações de diretório do objeto.
vcard:hasStreetAddress	Usado para suportar parâmetros de propriedade da propriedade de dados de endereço.
vcard:hasTelephone	Para especificar o número de telefone para comunicação telefônica com o objeto.
vcard:hasTitle	Usado para suportar parâmetros de propriedade para a propriedade de dados do título.
vcard:hasUID	Para especificar um valor que represente um identificador globalmente único correspondente ao objeto.
vcard:hasURL	Para especificar um URL associado ao objeto.
vcard:hasValue	Usado para indicar o valor do recurso de uma propriedade do objeto que requer parâmetros de propriedade.
vcard:key	-
vcard:logo	-
vcard:name	-
vcard:organization	Use a propriedade de dados do nome da organização ('organization name').
vcard:photo	-
vcard:sound	-
vcard:telephone	-
vcard:url	-

Fonte: adaptação e tradução de Iannella e McKinney (2014).

A seguir, no Quadro 28 constam as propriedades de dados, acompanhadas de suas descrições, quando houver, assim como informações sobre continuidade, com as propriedades obsoletas em cinza.

Quadro 28 – Propriedades de dados do vCard Ontology

Propriedade	Descrição
-------------	-----------

vcard:additional-name	O nome adicional associado ao objeto.
vcard:anniversary	A data do casamento, ou equivalente, do objeto.
vcard:birthday	Para especificar a data de nascimento do objeto.
vcard:category	As informações da categoria sobre o objeto, também conhecidas como etiquetas (<i>tags</i>).
vcard:class	Propriedade obsoleta.
vcard:country-name	O nome do país associado ao endereço do objeto.
vcard:extended-address	Propriedade obsoleta.
vcard:family-name	O nome da família associado ao objeto.
vcard:formatted-name	O texto formatado correspondente ao nome do objeto.
vcard:given-name	O nome dado associado ao objeto.
vcard:honorific-prefix	O prefixo honorífico do nome associado ao objeto.
vcard:honorific-suffix	O sufixo honorífico do nome associado ao objeto.
vcard:label	Propriedade obsoleta.
vcard:language	Para especificar o idioma que pode ser usado para entrar em contato com o objeto. Também pode ser usado como um parâmetro de propriedade.
vcard:latitude	Propriedade obsoleta. Ver 'hasGeo'.
vcard:locality	A localidade (por exemplo, cidade ou vila) associada ao endereço do objeto.
vcard:longitude	Propriedade obsoleta. Ver 'hasGeo'.
vcard:mailer	Propriedade obsoleta.
vcard:nickname	O apelido associado ao objeto.
vcard:note	Uma nota associada ao objeto.
vcard:organization-name	Para especificar o nome da organização associado ao objeto.
vcard:organizational-unit	Para especificar o nome da unidade organizacional associado ao objeto.
vcard:post-office-box	Propriedade obsoleta.
vcard:postal-code	O código postal associado ao endereço do objeto.
vcard:prodid	Para especificar o identificador para o produto que criou o objeto.
vcard:region	A região (por exemplo, estado ou província) associada ao endereço do objeto.
vcard:revision	Para especificar informações de revisão sobre o objeto.
vcard:role	Para especificar a função ou parte desempenhada em uma situação específica pelo objeto.
vcard:sort-string	Para especificar a sequência a ser usada para classificação específica do idioma nacional. Usado apenas como um parâmetro de propriedade.
vcard:street-address	O endereço da rua associado ao endereço do objeto.
vcard:tz	Para indicar informações de fuso horário específicas ao objeto. Também pode ser usado como um parâmetro de propriedade.
vcard:title	Para especificar a posição ou emprego do objeto.
vcard:value	Usado para indicar o valor literal de uma propriedade de dados que requer parâmetros de propriedade.

Fonte: adaptação e tradução de Iannella e McKinney (2014).

Na Figura 13 abaixo apresenta-se um exemplo de descrição em *vCard Ontology*, em RDF/XML, contendo informações de apelido, número de telefone,

email, endereço e localização geográfica sobre um indivíduo chamado ‘Corky Crystal’.

Figura 13 – Exemplo de uso do vCard Ontology

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:vcard="http://www.w3.org/2006/vcard/ns#">
  <vcard:Individual rdf:about="http://example.com/me/corky">
    <vcard:fn>Corky Crystal</vcard:fn>
    <vcard:nickname>Corks</vcard:nickname>
    <vcard:hasTelephone rdf:parseType="Resource">
      <vcard:hasValue rdf:resource="tel:+617555555555"/>
      <rdf:type rdf:resource="http://www.w3.org/2006/vcard/ns#Home"/>
      <rdf:type rdf:resource="http://www.w3.org/2006/vcard/ns#Voice"/>
    </vcard:hasTelephone>
    <vcard:hasEmail rdf:resource="mailto:corky@example.com"/>
    <vcard:hasAddress rdf:parseType="Resource">
      <vcard:street-address>111 Lake Drive</vcard:street-address>
      <vcard:locality>WonderCity</vcard:locality>
      <vcard:postal-code>5555</vcard:postal-code>
      <vcard:country-name>Australia</vcard:country-name>
      <rdf:type rdf:resource="http://www.w3.org/2006/vcard/ns#Home"/>
    </vcard:hasAddress>
  </vcard:Individual>
</rdf:RDF>
```

Fonte: Iannella e McKinney (2014).

A partir da adoção das classes e propriedades do *vCard Ontology* busca-se proporcionar a representação de informações relacionados a pessoas, geralmente presentes em agendas, como endereços, nomes, números de telefone e *e-mails*, além de permitir a categorização dos indivíduos de acordo com suas relações (subclasses de ‘RelationType’) (IANNELLA; MCKINNEY, 2014).

5.4 PROV Ontology (PROV-O)

O PROV Ontology (PROV-O) faz parte da chamada família de documentos PROV, decorrente dos esforços do *Provenance Working Group* do W3C, buscando viabilizar a representação e intercâmbio de informações de proveniência geradas em sistemas variados (ARAKAKI, 2019; LEBO; SAHOO; MCGUINNESS, 2013). A família PROV é formada por: PROV *Data Model* (PROV-DM), *Provenance Notation* (PROV-N), *Constraints of the PROV Data Model* (PROV-CONSTRAINTS) e PROV *Ontology* (PROV-O).

O PROV-DM é um “[...] modelo conceitual de dados que forma uma base para a família de especificações W3C de proveniência (PROV).” (MOREAU; MISSIER, 2013a, n.p., tradução nossa), proporcionando interoperabilidade de proveniência entre diferentes sistemas. De acordo com Arakaki (2019, p. 49), o modelo fornece “[...] estruturas centrais formando a essência de informações de procedência.”.

Já o PROV-N é voltado para a compreensão humana, buscando disponibilizar notações para proveniência, sendo que “[...] facilita o mapeamento do modelo de dados PROV para sintaxe concreta e é usado como base para uma semântica formal do PROV.” (MOREAU; MISSIER, 2013b, n.p., tradução nossa).

O PROV-CONSTRAINTS dispõe as restrições ao PROV-DM, definindo um subconjunto de instâncias – as chamadas instâncias PROV válidas – para “[...] satisfazer certas definições, inferências e restrições.” (CHENEY; MISSIER; MOREAU, 2013, n.p., tradução nossa) de outros padrões da *Web*.

O PROV-O “[...] expressa o modelo de dados PROV (PROV-DM) usando a OWL2 Web Ontology Language (OWL2).” (LEBO; SAHOO; MCGUINNESS, 2013, n.p., tradução nossa), permitindo que seja realizado o mapeamento do PROV-DM para RDF. De acordo com Lebo, Sahoo e McGuinness (2013, n.p., tradução nossa), o PROV-O

[...] fornece um conjunto de classes, propriedades e restrições que podem ser usadas para representar e trocar informações de proveniência geradas em diferentes sistemas e em diferentes contextos. Também pode ser especializado para criar novas classes e propriedades para modelar informações de proveniência para diferentes aplicações e domínios.

Os termos do PROV-O são classificados em três categorias: termos de ponto de partida (*starting point terms*), termos expandidos (*expanded terms*) e termos qualificados (*qualified terms*) (LEBO; SAHOO; MCGUINNESS, 2013).

Os termos de ponto de partida proveém um conjunto básico de classes e propriedades, servindo de base para o PROV-O. Fazem parte dessa categoria três classes e nove propriedades. No Quadro 29 a seguir apresenta-se as classes.

Quadro 29 – Classes de ponto de partida do PROV *Ontology* (PROV-O)

Classe	Descrição
prov:Entity	Uma coisa física, digital, conceitual ou outro tipo de coisa com alguns aspectos fixos; entidades podem ser reais ou imaginárias.
prov:Activity	Algo que ocorre durante um período de tempo e atua sobre ou com entidades; pode incluir consumir, processar, transformar, modificar, realocar, usar ou gerar entidades.
prov:Agent	Algo que tenha alguma forma de responsabilidade por uma atividade que está ocorrendo, pela existência de uma entidade ou pela atividade de outro agente.

Fonte: adaptação e tradução de Lebo, Sahoo e McGuinness (2013).

No Quadro 30 abaixo estão presentes as propriedades de ponto de partida.

Quadro 30 – Propriedades de ponto de partida do PROV *Ontology* (PROV-O)

Propriedade	Descrição
prov:wasGeneratedBy	Indica o que gerou a entidade. Geração é a conclusão da produção de uma nova entidade por uma atividade.
prov:wasDerivedFrom	Indica a entidade preexistente da qual uma derivação foi originada. Uma derivação é uma transformação de uma entidade em outra, uma atualização de uma entidade resultando em uma nova ou a construção de uma nova entidade com base em uma entidade preexistente.
prov:wasAttributedTo	Indica a qual agente uma entidade foi atribuída. Atribuição é a designação de uma entidade a um agente.
prov:startedAtTime	A hora em que uma atividade teve início. Início é quando uma atividade é considerada iniciada por uma entidade, conhecida como gatilho. Qualquer uso, geração ou invalidação envolvendo uma atividade segue o início da atividade.
prov:used	Indica uma entidade que foi usada por uma atividade. Uso é o começo da utilização de uma entidade por uma atividade.
prov:wasInformedBy	Indica uma atividade a1 que informou uma atividade a2. Comunicação é a troca de uma entidade por duas atividades, uma atividade usando a entidade gerada pela outra.
prov:endedAtTime	A hora em que uma atividade teve fim. Fim é quando uma atividade é considerada finalizada por uma entidade, conhecida como gatilho. Qualquer uso, geração ou invalidação envolvendo uma atividade precede o fim da atividade.
prov:wasAssociatedWith	Indica o agente responsável pela ocorrência de uma atividade. Uma associação de atividades é uma atribuição de responsabilidade a um agente para uma atividade, indicando que o agente teve uma função na atividade. Além disso, permite que um plano seja especificado, que é o plano pretendido pelo agente para atingir alguns objetivos no contexto desta atividade.
prov:actedOnBehalfOf	Indica o agente ao qual foi delegada responsabilidade. Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica como delegado ou representante, enquanto o agente que ele atua em nome de mantém a responsabilidade pelo resultado do trabalho delegado. Por exemplo, um aluno agiu em nome de seu supervisor, que agiu em nome do presidente do departamento, que agiu em nome da universidade; todos esses agentes são responsáveis de alguma forma pela atividade que ocorreu, mas não dizemos explicitamente quem é o responsável e em que grau.

Fonte: adaptação e tradução de Lebo, Sahoo e Mccguinness (2013).

Esses termos, conforme colocam Lebo, Sahoo e Mccguinness (2013, n.p., tradução nossa), “[...] são usados para criar descrições simples de proveniência que podem ser elaboradas usando termos de outras categorias.”.

Na categoria de termos expandidos constam sete classes e 16 propriedades. As classes expandidas estão presentes no Quadro 31.

Quadro 31 – Classes expandidas do PROV *Ontology* (PROV-O)

Classe	Descrição
prov:Collection	Uma coleção é uma entidade que fornece uma estrutura para alguns constituintes, que também são entidades.
prov:EmptyCollection	Uma coleção vazia é uma coleção sem membros.
prov:Bundle	Um pacote é um conjunto nomeado de descrições de proveniência e é ele próprio uma Entidade, permitindo que a proveniência de proveniência seja expressa.
prov:Person	Agentes pessoais são pessoas.
prov:SoftwareAgent	Um agente de software é o <i>software</i> em execução.
prov:Organization	Uma organização é uma instituição social ou jurídica, como empresa, sociedade etc.
prov:Location	Um local pode ser um local geográfico identificável (ISO 19112), mas também pode ser um local não geográfico, como um diretório, linha ou coluna. Como tal, existem inúmeras maneiras pelas quais o local pode ser expresso, como por uma coordenada, endereço, ponto de referência etc.

Fonte: adaptação e tradução de Lebo, Sahoo e Mccguinness (2013).

No Quadro 32 abaixo estão descritas as propriedades expandidas do PROV-O.

Quadro 32 – Propriedades expandidas do PROV *Ontology* (PROV-O)

Propriedade	Descrição
prov:alternateOf	Duas entidades alternativas apresentam aspectos da mesma coisa. Esses aspectos podem ser iguais ou diferentes, e as entidades alternativas podem ou não se sobrepor no tempo.
prov:specializationOf	Uma entidade que é uma especialização de outra compartilha todos os aspectos desta e, adicionalmente, apresenta aspectos mais específicos da mesma coisa que a última. Exemplos de aspectos incluem um período de tempo, uma abstração e um contexto associado à entidade.
prov:generatedAtTime	A hora em que uma entidade foi completamente criada e está disponível para uso.
prov:hadPrimarySource	Indica a fonte primária de uma derivação. Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento do estudo do tópico, sem se beneficiar em retrospectiva. Como tal, é importante que as fontes secundárias façam referência às fontes primárias das quais elas derivaram, para que sua confiabilidade possa ser investigada.
prov:value	Fornecer um valor que é uma representação direta de uma entidade.

prov:wasQuotedFrom	Indica a entidade da qual derivou a citação. Uma citação é a repetição (de parte ou da totalidade) de uma entidade, como texto ou imagem, por alguém que pode ou não ser o autor original. A citação é um caso particular de derivação.
prov:wasRevisionOf	Indica a entidade da qual derivou a revisão. Uma revisão é uma derivação para a qual a entidade resultante é uma versão revisada de algum original. A implicação aqui é que a entidade resultante contém conteúdo substancial do original. A revisão é um caso particular de derivação.
prov:invalidatedAtTime	A hora em que uma entidade foi invalidada. Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade. Qualquer geração ou uso de uma entidade precede sua invalidação.
prov:wasInvalidatedBy	Indica o que foi responsável pela invalidação de uma entidade.
prov:hadMember	Uma coleção é uma entidade que fornece uma estrutura para alguns constituintes, que também são entidades. Diz-se que esses constituintes são membros das coleções.
prov:wasStartedBy	Indica a entidade que deu início a uma atividade.
prov:wasEndedBy	Indica a entidade que finalizou uma atividade.
prov:invalidated	Indica a entidade que foi invalidada.
prov:influenced	Influência é a capacidade de uma entidade, atividade ou agente de afetar o caráter, desenvolvimento ou comportamento de outro por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
prov:atLocation	A localização de qualquer recurso.
prov:generated	Indica a entidade que foi gerada.

Fonte: adaptação e tradução de Lebo, Sahoo e Mccguinness (2013).

Essas classes e propriedades expandidas podem ser utilizadas para relacionar classes dos termos de ponto de partida, sendo que muitas delas são subclasses e subpropriedades dos termos de ponto de partida (LEBO; SAHOO; MCGUINNESS, 2013).

Na categoria de termos qualificados existem 20 classes e 25 propriedades. As classes qualificadas estão dispostas no Quadro 33 a seguir.

Quadro 33 – Classes qualificadas do PROV *Ontology* (PROV-O)

Classe	Descrição
prov:Influence	Influência é a capacidade de uma entidade, atividade ou agente de afetar o caráter, desenvolvimento ou comportamento de outro por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
prov:EntityInfluence	A capacidade de influência de uma entidade.
prov:Usage	Uso é o começo da utilização de uma entidade por uma atividade.
prov:Start	Início é quando uma atividade é considerada iniciada por uma entidade, conhecida como gatilho.

prov:End	Fim é quando uma atividade é considerada finalizada por uma entidade, conhecida como gatilho.
prov:Derivation	Uma derivação é uma transformação de uma entidade em outra, uma atualização de uma entidade resultando em uma nova ou a construção de uma nova entidade com base em uma entidade preexistente.
prov:PrimarySource	Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento do estudo do tópico, sem se beneficiar em retrospectiva.
prov:Quotation	Uma citação é a repetição (de parte ou da totalidade) de uma entidade, como texto ou imagem, por alguém que pode ou não ser o autor original.
prov:Revision	Uma revisão é uma derivação para a qual a entidade resultante é uma versão revisada de algum original.
prov:ActivityInfluence	A capacidade de influência de uma atividade.
prov:Generation	Geração é a conclusão da produção de uma nova entidade por uma atividade.
prov:Communication	Comunicação é a troca de uma entidade por duas atividades, uma atividade usando a entidade gerada pela outra.
prov:Invalidation	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade.
prov:AgentInfluence	A capacidade de influência de um agente.
prov:Attribution	Atribuição é a designação de uma entidade a um agente.
prov:Association	Uma associação de atividades é uma atribuição de responsabilidade a um agente para uma atividade, indicando que o agente teve uma função na atividade.
prov:Plan	Um plano é uma entidade que representa um conjunto de ações ou etapas pretendidas por um ou mais agentes para atingir alguns objetivos.
prov:Delegation	Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica como delegado ou representante, enquanto o agente que ele atua em nome de mantém a responsabilidade pelo resultado do trabalho delegado.
prov:InstantaneousEvent	Um evento instantâneo, ou evento abreviado, acontece no mundo e marca uma mudança no mundo, em suas atividades e em suas entidades. Os eventos incluem geração, uso ou invalidação de entidades, bem como início ou término de atividades.
prov:Role	O papel é a função de uma entidade ou agente em relação a uma atividade, no contexto de uso, geração, invalidação, associação, início e fim.

Fonte: adaptação e tradução de Lebo, Sahoo e McGuinness (2013).

No Quadro 34 abaixo encontram-se dispostas as propriedades qualificadas do PROV-O.

Quadro 34 – Propriedades qualificadas do PROV *Ontology* (PROV-O)

Propriedade	Descrição
prov:wasInfluencedBy	Indica entidade, atividade ou agente responsável por

Propriedade	Descrição
	influenciar outro.
prov:qualifiedInfluence	Influência é a capacidade de uma entidade, atividade ou agente de afetar o caráter, desenvolvimento ou comportamento de outro por meio de uso, início, fim, geração, invalidação, comunicação, derivação, atribuição, associação ou delegação.
prov:qualifiedGeneration	Geração é a conclusão da produção de uma nova entidade por uma atividade.
prov:qualifiedDerivation	Uma derivação é uma transformação de uma entidade em outra, uma atualização de uma entidade resultando em uma nova ou a construção de uma nova entidade com base em uma entidade preexistente.
prov:qualifiedPrimarySource	Uma fonte primária para um tópico refere-se a algo produzido por algum agente com experiência direta e conhecimento sobre o tópico, no momento do estudo do tópico, sem se beneficiar em retrospectiva.
prov:qualifiedQuotation	Uma citação é a repetição (de parte ou da totalidade) de uma entidade, como texto ou imagem, por alguém que pode ou não ser o autor original.
prov:qualifiedRevision	Uma revisão é uma derivação para a qual a entidade resultante é uma versão revisada de algum original.
prov:qualifiedAttribution	Atribuição é a designação de uma entidade a um agente.
prov:qualifiedInvalidation	Invalidação é o início da destruição, cessação ou expiração de uma entidade existente por uma atividade.
prov:qualifiedStart	Início é quando uma atividade é considerada iniciada por uma entidade, conhecida como gatilho.
prov:qualifiedUsage	Uso é o começo da utilização de uma entidade por uma atividade.
prov:qualifiedCommunication	Comunicação é a troca de uma entidade por duas atividades, uma atividade usando a entidade gerada pela outra.
prov:qualifiedAssociation	Uma associação de atividades é uma atribuição de responsabilidade a um agente para uma atividade, indicando que o agente teve uma função na atividade.
prov:qualifiedEnd	Fim é quando uma atividade é considerada finalizada por uma entidade, conhecida como gatilho.
prov:qualifiedDelegation	Delegação é a atribuição de autoridade e responsabilidade a um agente (por si ou por outro agente) para realizar uma atividade específica como delegado ou representante, enquanto o agente que ele atua em nome de mantém a responsabilidade pelo resultado do trabalho delegado.
prov:influencer	Essa propriedade é usada como parte do padrão de influência qualificado. Subclasses de prov:Influence usam essas subpropriedades para referenciar o recurso (Entidade, Agente ou Atividade) cuja influência está sendo qualificada.
prov:entity	Essa propriedade (prov:entity) faz referência a uma entidade (prov:Entity) que influenciou um recurso. Ela se aplica a uma influência de entidade (prov:EntityInfluence), fornecida por uma subpropriedade de prov:qualifiedInfluence da entidade

Propriedade	Descrição
	(prov:Entity), atividade (prov:Activity) ou agente (prov:Agent) influenciado.
prov:hadUsage	O uso opcional envolvido na derivação de uma entidade.
prov:hadGeneration	A geração opcional envolvida na derivação de uma entidade.
prov:activity	Essa propriedade (prov:activity) faz referência a uma atividade (prov:Activity) que influenciou um recurso. Ela se aplica a uma influência de atividade (prov:ActivityInfluence), fornecida por uma subpropriedade de prov:qualifiedInfluence da entidade (prov:Entity), atividade (prov:Activity) ou agente (prov:Agent) influenciado.
prov:agent	Essa propriedade (prov:agent) faz referência a um agente (prov:Agent) que influenciou um recurso. Ela se aplica a uma influência de agente (prov:AgentInfluence), fornecida por uma subpropriedade de prov:qualifiedInfluence da entidade (prov:Entity), atividade (prov:Activity) ou agente (prov:Agent) influenciado.
prov:hadPlan	O plano opcional adotado por um agente em associação com alguma atividade
prov:hadActivity	A atividade opcional de uma influência, que usou, gerou, invalidou ou era de responsabilidade de alguma entidade
prov:atTime	A hora em que um evento instantâneo (prov:InstantaneousEvent) ocorreu, no formato xsd:dateTime.
prov:hadRole	A função opcional que uma entidade assumiu no contexto de uma atividade.

Fonte: adaptação e tradução de Lebo, Sahoo e McGuinness (2013).

As classes e propriedades qualificadas tem como objetivo fornecer “[...] informações elaboradas sobre relações binárias declaradas usando propriedades de ponto de partida e expandidas.” (LEBO; SAHOO; MCGUINNESS, 2013, n.p., tradução nossa), de modo a acrescentar maiores detalhes a esses termos.

5.5 Vocabulary of Interlinked Datasets (VOID)

De acordo com o *World Wide Web Consortium* (2011), o VOID, um vocabulário RDF *Schema*, foi desenvolvido para o estabelecimento de metadados para conjuntos de dados RDF, de modo a promover e auxiliar em ações como a descoberta de dados, catalogação e arquivamento de conjuntos de dados, auxiliando as atividades de busca e recuperação de dados dos usuários.

De modo geral, o VOID cobre quatro áreas de metadados: metadados gerais (*general metadata*), metadados de acesso (*access metadata*), metadados estruturais

(*structural metadata*) e metadados para descrição de ligações entre conjuntos de dados (*description of links between datasets*).

Os metadados gerais são os metadados utilizados para descrever as características gerais dos conjuntos de dados, como título, descrição, licença e assunto (WORLD WIDE WEB CONSORTIUM, 2011). Na Figura 14 é exibido um exemplo de uso de metadados gerais do Void.

Figura 14 – Exemplo de uso de metadados gerais do *Vocabulary of Interlinked Datasets* (VoID)

```

:DBpedia a void:Dataset;
  dcterms:title "DBpedia";
  dcterms:description "RDF data extracted from Wikipedia";
  dcterms:contributor :FU_Berlin;
  dcterms:contributor :University_Leipzig;
  dcterms:contributor :OpenLink_Software;
  dcterms:contributor :DBpedia_community;
  dcterms:source <http://dbpedia.org/resource/Wikipedia>;
  dcterms:modified "2008-11-17"^^xsd:date;
.

:FU_Berlin a foaf:Organization;
  rdfs:label "Freie Universität Berlin";
  foaf:homepage <http://www.fu-berlin.de/>;
.

# Similar descriptions of the other contributors go here

```

Fonte: *World Wide Web Consortium* (2011).

O exemplo acima apresenta a descrição geral do DBpedia¹⁵ com a utilização de elementos do DC *Terms*, RDFS e FOAF.

Os metadados de acesso são os metadados “[...] utilizados para descrever métodos de acesso às triplas RDF.” (WORLD WIDE WEB CONSORTIUM, 2011, n.p., tradução nossa). Na Figura 15 é possível observar um exemplo de uso:

Figura 15 – Exemplo de uso de metadados de acesso do *Vocabulary of Interlinked Datasets* (VoID)

```

:NYTimes a void:Dataset;
  void:dataDump <http://data.nytimes.com/people.rdf>;
  void:dataDump <http://data.nytimes.com/organizations.rdf>;
  void:dataDump <http://data.nytimes.com/locations.rdf>;
  void:dataDump <http://data.nytimes.com/descriptors.rdf>;
.

```

Fonte: *World Wide Web Consortium* (2011).

¹⁵ “DBpedia é um esforço comunitário de origem coletiva para extrair conteúdo estruturado das informações criadas em vários projetos da Wikimedia.” (DBPEDIA, c2019, n.p., tradução nossa).

De acordo com o exemplo da Figura 14, o conjunto de dados *NYTimes* está disponível como um conjunto de quatro arquivos RDF.

Os metadados estruturais são os metadados utilizados para fornecer informações relacionadas ao esquema utilizado pelo conjunto de dados, bem como sobre sua estrutura interna, sendo informações sobre vocabulários, tamanho e exemplos de recursos. A Figura 16 exemplifica a utilização de metadados estruturais.

Figura 16 – Exemplo de uso de metadados estruturais do *Vocabulary of Interlinked Datasets* (VOID)

```
:LiveJournal a void:Dataset;  
  void:vocabulary <http://xmlns.com/foaf/0.1/>;  
 .
```

Fonte: *World Wide Web Consortium* (2011).

O exemplo da Figura 16 acima diz que o conjunto de dados *LiveJournal* utiliza o FOAF *Vocabulary*.

Os metadados para descrição de ligações entre conjuntos de dados são os metadados utilizados para descrever *linksets*, em que um *linkset* corresponde a uma coleção de ligações RDF entre conjuntos de dados. “É um conjunto de triplas RDF em que todos os assuntos estão em um conjunto de dados e todos os objetos em outro conjunto de dados.” (WORLD WIDE WEB CONSORTIUM, 2011, n.p., tradução nossa). Na Figura 17 abaixo observa-se um exemplo de metadados para *linksets*.

Figura 17 – Exemplo de uso de metadados para descrição de *linksets* do *Vocabulary of Interlinked Datasets* (VOID)

```
:DBpedia_Geonames a void:Linkset;  
  void:target :DBpedia;  
  void:target :Geonames;  
  void:subset :DBpedia;  
  void:triples 252000;  
 .
```

Fonte: *World Wide Web Consortium* (2011).

De acordo com o exemplo, o conjunto de dados *DBpedia* possui um *linkset* de 252.000 ligações ao *GeoNames*¹⁶.

Observa-se que outros vocabulários são utilizados em conjunto com o VoID, como o FOAF *Vocabulary* e DC *Terms*. O VoID em si possui quatro classes e 27 propriedades (CYGANIAK *et al.*, 2011). As classes encontram-se listadas no Quadro 35 abaixo.

Quadro 35 – Classes do *Vocabulary of Interlinked Datasets* (VoID)

Classe	Descrição	Superclasse	Subclasse
void:Dataset	Um conjunto de triplas RDF que são publicadas, mantidas ou agregadas por um único provedor.	dctype:Dataset	void:Linkset
void:DatasetDescription	Um recurso da Web cujo tópico primário (foaf:primaryTopic) ou tópicos (foaf:topics) incluem conjuntos de dados (void:Datasets).	foaf:Document	-
void:Linkset	Uma coleção de ligações RDF entre dois conjuntos de dados (void:Datasets).	void:Dataset	-
void:TechnicalFeature	Um recurso técnico de um conjunto de dados (void:Dataset), como um formato de serialização RDF suportado.	-	-

Fonte: adaptação e tradução de Cyganiak *et al.* (2011).

As propriedades do VoID encontram-se descritas no Quadro 36 a seguir.

Quadro 36 – Propriedades do *Vocabulary of Interlinked Datasets* (VoID)

Propriedade	Descrição	Domínio
void:class	A classe (rdfs:Class) que é o tipo (rdf:type) de todas as entidades em uma partição baseada em classe.	void:Dataset
void:classPartition	Um subconjunto de um conjunto de dados (void:Dataset) que contém apenas as entidades de uma determinada classe (rdfs:Class).	
void:classes	O número total de classes distintas em um conjunto de dados (void:Dataset). Em outras palavras, o número de recursos distintos que ocorrem como objetos de tipos (rdf:type) de triplas no	

¹⁶ O *GeoNames* é uma base de dados gratuita que disponibiliza “[...] nomes de lugares em vários idiomas, altitude, população e outros de várias fontes.” (GEONAMES, [2020?], n.p., tradução nossa).

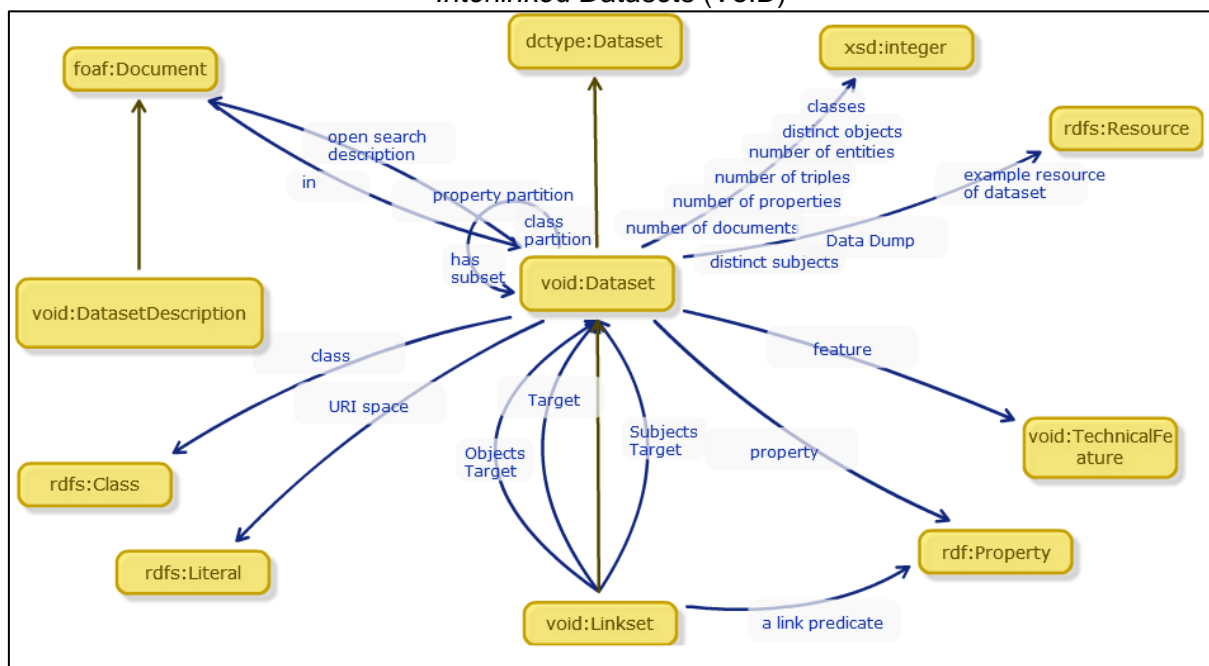
Propriedade	Descrição	Domínio
	conjunto de dados.	
void:dataDump	Um despejo RDF, parcial ou completo, de um conjunto de dados (void:Dataset).	
void:distinctObjects	O número total de objetos distintos em um conjunto de dados (void:Dataset). Em outras palavras, o número de recursos distintos que ocorrem na posição do objeto de triplas no conjunto de dados. Os literais estão incluídos nesta contagem.	
void:distinctSubjects	O número total de assuntos distintos em um conjunto de dados (void:Dataset). Em outras palavras, o número de recursos distintos que ocorrem na posição de sujeito de triplas no conjunto de dados.	
void:documents	O número total de documentos, para conjuntos de dados publicados como um conjunto de documentos individuais, como documentos RDF/XML ou páginas da <i>Web</i> anotadas por RDFa. Documentos não RDF, como páginas da <i>Web</i> em HTML ou imagens, geralmente não são incluídos nessa contagem. Essa propriedade é destinada a conjuntos de dados em que é difícil determinar o número total de triplas ou entidades. Triplas (void:triples) e entidades (void:entities) devem ser preferidas sempre que possível.	
void:entities	O número total de entidades que são descritas em um conjunto de dados (void:Dataset).	
void:exampleResource	Recurso de exemplo do conjunto de dados.	
void:feature	Característica.	
void:openSearchDescription	Um documento de descrição <i>OpenSearch</i> para um serviço de pesquisa de texto livre em um conjunto de dados (void:Dataset).	
void:properties	O número total de propriedades distintas em um conjunto de dados (void:Dataset). Em outras palavras, o número de recursos distintos que ocorrem na posição predicado de triplas no conjunto de dados.	
void:property	A propriedade (rdf:Property) que é o predicado de todas as triplas em uma partição baseada em propriedade.	
void:propertyPartition	Um subconjunto de um conjunto de dados (void:Dataset) que contém apenas as triplas de uma determinada propriedade (rdf:Property).	

Propriedade	Descrição	Domínio
void:rootResource	Um conceito ou ponto de entrada principal para um conjunto de dados (void:Dataset) estruturado de maneira semelhante a uma árvore. Todos os recursos em um conjunto de dados podem ser acessados seguindo as ligações dos recursos raiz em um pequeno número de etapas.	
void:sparqlEndpoint	Indica um SPARQL <i>endpoint</i> .	
void:subset	Indica um subconjunto.	
void:triples	O número total de triplas contidas em um conjunto de dados (void:Dataset).	
void:uriLookupEndpoint	Define um protocolo simples de pesquisa de URI para acessar um conjunto de dados.	
void:uriRegexPattern	Define um padrão de expressão regular que corresponde a URIs no conjunto de dados.	
void:uriSpace	Um URI que é um prefixo de sequência comum de todos os URIs da entidade em um conjunto de dados (void:Dataset).	
void:vocabulary	Um vocabulário usado no conjunto de dados.	
void:linkPredicate	Um predicado de ligação.	
void:objectsTarget	O conjunto de dados que descreve os objetos das triplas contidas no <i>Linkset</i> .	void:Linkset
void:subjectsTarget	O conjunto de dados que descreve os sujeitos de triplas contidas no <i>Linkset</i> .	
void:target	Um dos dois conjuntos de dados vinculados pelo <i>Linkset</i> .	
void:inDataset	Aponta para o conjunto de dados (void:Dataset) do qual um documento faz parte.	foaf:Document

Fonte: adaptação e tradução de Cyganiak *et al.* (2011).

Nota-se que a propriedade 'void:inDataset' possui como domínio um elemento advindo do FOAF *Vocabulary*. Além disso, as propriedades e classes possuem relações a outros vocabulários além do FOAF, como pode ser melhor observado pelo diagrama apresentado na Figura 18 a seguir.

Figura 18 – Diagrama das relações entre as classes e propriedades do *Vocabulary of Interlinked Datasets (VOID)*



Fonte: Cyganiak *et al.* (2011).

De acordo com Akar *et al.* (2012, n.p., tradução nossa), “Como o VOID se baseia na alma da web de dado baseada em grafo, ele fornece uma forma forte de descrever metadados que permite a descoberta de conjuntos de dados nos quais as consultas são distribuídas.”. Com a utilização do Void torna-se possível a descoberta e a seleção de conjuntos de dados relevantes e, até mesmo, de forma automatizada (AKAR *et al.*, 2012).

Em seu estudo, os autores abordam a ferramenta *Web of Data Query Analyzer* (WoDQA), desenvolvida para atuar sobre conjuntos de dados descritos em Void: “O WoDQA visa consultar a web de conjuntos de dados interligados¹⁷ usando conjuntos de dados VOID e descrições de linksets para decidir conjuntos de dados relevantes para uma consulta.” (AKAR *et al.*, 2012, n.p., tradução nossa). Akar *et al.* (2012) relatam que não foi possível averiguar compreensivamente a eficácia do WoDQA, uma vez que os metadados de SPARQL *endpoint*, descrições de *linksets* e vocabulários encontraram-se ausentes na maioria dos conjuntos de dados em Void.

¹⁷ Akar *et al.* (2012) utilizam a expressão “web de conjuntos de dados interligados (*web of interlinked datasets*)” para se referirem à LOD *Cloud*.

6 UTILIZAÇÃO DO DATA CATALOG VOCABULARY (DCAT)

Pensando-se em ter uma compreensão da adoção do DCAT, realizou-se a verificação de uso do vocabulário em plataformas registradas no re3data, o *Registry of Research Data Repositories*.

Re3data é um registro global de repositórios de dados de pesquisa que cobre repositórios de dados de pesquisa de diferentes disciplinas acadêmicas. Inclui repositórios que permitem o armazenamento permanente e o acesso a conjuntos de dados para pesquisadores, órgãos de financiamento, editores e instituições acadêmicas. (REGISTRY OF RESEARCH DATA REPOSITORIES, [2020?], n.p., tradução nossa).

Mantidos pelo DataCite desde 2015, a plataforma possui como parceiros: a *Berlin School of Library and Information Science*, da Humboldt-Universität zu Berlin; o *Helmholtz Open Science Office*, do *GFZ German Research Centre for Geosciences*; a *KIT Library*, do *Karlsruhe Institute of Technology (KIT)*; e as bibliotecas da *Purdue University* (REGISTRY OF RESEARCH DATA REPOSITORIES, [2020?]).

No *site* do re3data foi realizada a busca pelas expressões “Data Catalog Vocabulary” e “DCAT”. Após a busca, selecionou-se no filtro ‘*Metadata standards*’ o vocabulário, presente como “DCAT – Data Catalog Vocabulary”, conforme a Figura 19 a seguir.

Figura 19 – Busca pelo *Data Catalog Vocabulary*, realizada no re3data

The screenshot shows the re3data search interface. The search bar at the top contains the text "data catalog vocabulary". Below the search bar, there are navigation buttons: "← Previous", "1", and "Next →". The results section indicates "Found 19 result(s)". Three results are visible:

- Open Data by Socrata**: Socrata. Content type(s): Audiovisual data, Networkbased data, Standard office documents, Plain text, Raw data, Images. Country: United States. Description: Socrata's cloud-based solution allows government organizations to put their data online, make data-driven decisions, operate more efficiently, and...
- JRC Data Catalogue**: EC's Joint Research Centre Data Catalogue. Subject(s): Economics, Life Sciences, Natural Sciences, Engineering Sciences, Social and Behavioural Sciences. Content type(s): Scientific and statistical data formats, Networkbased data, Databases, Images. Country: European Union. Description: The JRC Data Catalogue gives access to the multidisciplinary data produced and maintained by the Joint Research Centre, the European Commission, and provides advice and support to policies of the European Union.
- Open Government Canada - Open Data**: Gouvernement Ouvert Canada - données ouvertes. Subject(s): Humanities and Social Sciences, Life Sciences, Natural Sciences, Engineering Sciences. Content type(s): Standard office documents, Structured graphics, Scientific and statistical data formats, Raw data.

On the left side, there is a "Filter" sidebar with various categories. Under "Metadata standards", "DCAT - Data Catalog Vocabulary (19)" is highlighted with a red box. Other standards listed include CF (Climate and Forecast) Metadata Conventions (1), DIF - Directory Interchange Format (2), DataCite Metadata Schema (3), Dublin Core (4), FGDC/CSDGM - Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata (1), and ISO 19115 (5).

Fonte: *Registry of Research Data Repositories*, 2020b.

A busca gerou 19 resultados: *Ag Data Commons*; *Arctic Permafrost Geospatial Centre*; *Biological and Chemical Oceanography Data Management Office*; *COEMS Open Data*; *Data.gov.au*; *Etsin Research Data Finder*; *European Data Portal*; *European Union Open Data Portal*; *Fairdata IDA Research Data Storage Service*; *GOVDATA*; *ICOS Carbon Portal*; *JRC Data Catalogue*; *Norwegian Polar Data Centre*; *Open Data by Socrata*; *Open Data Kingston*; *Open Government Canada - Open Data*; *RDPCIDAT*; *Répertoire International des Sources Musicales (RISM)*; e *Repositorio de datos del Ministerio de Educación del Perú*. Considerando a totalidade de 2606¹⁸ plataformas registradas no re3data, as que afirmam utilizar o DCAT correspondem à cerca de 0,73% da plataforma.

Em seguida os portais foram acessados a fim de identificar o uso do DCAT. Em um primeiro momento, 1) buscou-se pela presença do DCAT nos formatos legíveis por máquina dos registros dos conjuntos de dados ou dos próprios catálogos. Em casos em que não estavam disponibilizados, 2) tentou-se entrar em contato com as plataformas por meio de *e-mail* a fim de melhor esclarecer a situação de adoção do vocabulário.

A seguir são apresentadas em maiores detalhes as análises e informações de cada um dos portais.

6.1 Ag Data Commons

O *Ag Data Commons* é um catálogo e repositório aberto e governamental do Departamento de Agricultura dos Estados Unidos (*United States Department of Agriculture*), responsável pelo armazenamento e disseminação de dados de pesquisa voltados à comunidade agrícola, como dados relacionados a alimentos e agricultura. O repositório também afirma buscar disponibilizar seus dados segundo os princípios FAIR (UNITED STATES DEPARTMENT OF AGRICULTURE, [2020?]).

Em seus registros existe uma categoria chamada “outros acessos” (*other access*), em que é possível acessar os seus metadados nos formatos JSON e RDF/XML (Figura 20).

¹⁸ Número referente à data de 21 de dezembro de 2020.

Figura 20 – Registro do repositório Ag Data Commons

The screenshot displays the record page for 'USDA's Expanded Flavonoid Database for the Assessment of Dietary Intakes – September 2014'. On the left, there is a 'Filter By:' section with 'License' set to 'Creative Commons CCZero'. Below it, the 'Other Access' section provides information on dataset metadata availability in JSON, RDF, and via the DKAN API. The 'Social' section lists links for Twitter, LinkedIn, Reddit, Google+, and Facebook. The main content area features a title, a description of the database's update and content, and a list of associated files: 'Data Dictionary' (PDF, 199.19 KB), 'README - USDA's Expanded Flavonoid Database for the Assessment of Dietary Intakes Documentation and User Guide' (HTML), and 'FDB-EXP.accdb' (ZIP, 1.78 MB). Each file has an 'Explore Data' button and an MD5 hash.

Fonte: *United States Department of Agriculture*, 2019.

É possível verificar a presença do DCAT, porém, em todos os registros analisados, os arquivos apresentaram erro de validação, conforme a Figura 21 abaixo.

Figura 21 – Metadados em formato RDF de registro do repositório Ag Data Commons

```

<rdf:RDF>
  <dc:Dataset>
    <error>
      <_type>Validation Error</_type>
      <name_or_id>Missing Value</name_or_id>
    </error>
  </dc:Dataset>
  <dc:Distribution/>
  <foaf:Agent rdf:about="https://data.nal.usda.gov/publisher/n0">
    <foaf:name>DKAN</foaf:name>
    <foaf:homepage>https://data.nal.usda.gov</foaf:homepage>
    <dc:type rdf:resource="http://purl.org/adms/publishertype/NonProfitOrganisation"/>
  </foaf:Agent>
</rdf:RDF>

```

Fonte: *United States Department of Agriculture*, 2019.

Assim, apesar de se utilizar de elementos do DCAT, a não-validação dos metadados referentes ao conjunto de dados gera um impedimento para que os benefícios do vocabulário sejam efetivos.

6.2 Arctic Permafrost Geospatial Centre

O *Arctic Permafrost Geospatial Centre* é um catálogo mantido pela seção *Permafrost Research* do Instituto Alfred Wegener, Centro Helmholtz de Pesquisa Polar e Marinha (*Alfred Wegener Institute – AWI, Helmholtz Centre for Polar and Marine Research*). Ele se configura como aberto, sendo seu foco os dados geoespaciais do permafrost do Ártico (ARCTIC PERMAFROST GEOSPATIAL CENTRE, [2020?]).

Ao abrir um registro, sob “acesso a metadados” (*metadata access*) é possível visualizar os metadados em DCAT nos formatos RDF/XML, Turtle e JSON-LD, bem como em JSON (Figura 22).

Figura 22 – Registro do catálogo *Arctic Permafrost Geospatial Centre*

Historic Lake Drainage from Remote Sensing Datasets, 1955-2017, Western Arctic Coastal Plain, Northern Alaska (US)

Data Preview

Publisher

NSF Arctic Data Center
The Arctic Data Center is the primary data and software repository for the Arctic section of NSF Polar Programs. [read more](#)

Social

Twitter

Facebook

License

Dataset **Groups** **Activity Stream**

Historic Lake Drainage from Remote Sensing Datasets, 1955-2017, Western Arctic Coastal Plain, Northern Alaska (US)

Authors: Jones, Benjamin M.; Arp, Christopher D.; Grosse, Guido; Nitze, Ingmar; Lara, Mark; Whitman, Matthew; Farquharson, Louise; Kanevskiy, Mikhail; Parsekian, Andrew; Breen, Amy; Ohara, Nori; Rangel, Rodrigo; Hinkel, Kenneth

We identified all lakes larger than 10 hectares (ha) that drained completely or partially (greater than 25 %) between 1955 and 2017 using historical (original) USGS topographic maps and aerial photography (1955) and Landsat Imagery (circa 1975, circa 2000, and annually since 2000). For each lake drainage event, we inferred the drainage mechanism and categorized the drainage pathway based on known lake drainage mechanisms using interpretation of high-resolution remote sensing data and field observations.

Detailed information about the methods can be found in the data publication and the publication to which this dataset is a supplement.

Citation

In order to use these data, you must cite this data set with the following citation:
Benjamin Jones, Christopher Arp, Guido Grosse, Ingmar Nitze, Mark Lara, et al. 2019. Historic Lake Drainage on the Western Arctic Coastal Plain in Northern Alaska from Remote Sensing Datasets, 1955-2017. Arctic Data Center. doi:10.18739/A2DR2P85H.

Contact

[✉ Jones, Benjamin M.](#)

Metadata Access

- [DCAT in RDF/XML-Format](#)
- [DCAT in Turtle-Format](#)
- [DCAT in JSON-LD-Format](#)
- [APGC Dataset Metadata in JSON-Format](#)

Fonte: *Arctic Permafrost Geospatial Centre*, 2019.

Na Figura 23 a seguir encontra-se parte dos metadados do registro apresentado na Figura 22, no formato RDF/XML.

Figura 23 – Metadados em formato RDF/XML de registro do catálogo *Arctic Permafrost Geospatial Centre*

```

<rdf:RDF>
  <dc:Dataset rdf:about="https://apgc.awi.de/dataset/8fd16115-b144-40a4-988e-46c79ca40b2b">
    <dc:identifier>DOI:10.18739/A2DR2P85H</dc:identifier>
    <dc:keyword>Thermokarst Lake</dc:keyword>
    <dc:keyword>Lake Drainage</dc:keyword>
    <dc:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2020-03-06T12:52:52.752961</dc:issued>
    <dc:landingPage rdf:resource="https://doi.org/10.18739/A2DR2P85H">
    <owl:versionInfo>1.0</owl:versionInfo>
  </dc:Dataset>
  <dc:title>
    Historic Lake Drainage from Remote Sensing Datasets, 1955-2017, Western Arctic Coastal Plain, Northern Alaska (US)
  </dc:title>
  <dc:distribution>
    <dc:Distribution rdf:about="https://apgc.awi.de/dataset/8fd16115-b144-40a4-988e-46c79ca40b2b/resource/5ea0096f-7f8e-4a8d-94b5-8a67d813b627">
      <dc:mediaType>image/png</dc:mediaType>
      <dc:byteSize rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">878173.0</dc:byteSize>
      <dc:title>Preview</dc:title>
      <dc:accessURL rdf:resource="https://apgc.awi.de/dataset/8fd16115-b144-40a4-988e-46c79ca40b2b/resource/5ea0096f-7f8e-4a8d-94b5-8a67d813b627/download/preview_historic_lake_drainage.png">
      <dc:format>PNG</dc:format>
    </dc:Distribution>
  </dc:distribution>
  <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2020-11-02T08:52:40.734721</dc:modified>
  <dc:keyword>Permafrost</dc:keyword>

```

Fonte: *Arctic Permafrost Geospatial Centre*, 2019.

Pode-se observar a presença do DCAT, estando presentes elementos como `dc:Dataset`, `dc:keyword`, `dc:landingPage`, `dc:distribution`, `dc:mediaType`, `dc:byteSize` e `dc:accessURL`.

6.3 Biological and Chemical Oceanography Data Management Office

O *Biological and Chemical Oceanography Data Management Office* (BCO-DMO) foi fundado pela Fundação Nacional da Ciência, dos Estados Unidos (U.S. *National Science Foundation*), e está localizado no Instituto Oceanográfico de Woods Hole (*Woods Hole Oceanographic Institution*). Seu repositório é aberto, e seu foco são os dados referentes às Ciências da Terra (BIOLOGICAL AND CHEMICAL OCEANOGRAPHY DATA MANAGEMENT OFFICE, c2020; REGISTRY OF RESEARCH DATA REPOSITORIES, 2020a).

Ao final da página de cada registro do repositório existem as opções de visualizá-lo nos formatos ISO, RDF/XML, JSON e HTML, conforme ilustra a Figura 24.

Figura 24 – Registro do repositório *Biological and Schemical Oceanography Data Management Office*

Program: Dimensions of Biodiversity (Dimensions of Biodiversity) Ocean Carbon and Biogeochemistry (OCB) Historical Datasets (Historical)
Principal Investigator: Michael W. Lomas (Bigelow Laboratory for Ocean Sciences) Adam Martiny (University of California-Irvine, UC Irvine)
Contact: Adam Martiny (University of California-Irvine, UC Irvine)
BCO-DMO Data Manager: Nancy Copley Stephen R. Gegg (Woods Hole Oceanographic Institution, WHOI BCO-DMO)
Version: 2
Version Date: 2020-12-21
Restricted: No
Validated: No
Current State: Under revision
Data URL: https://www.bco-dmo.org/dataset/526747/data
Global ocean Particulate Organic Matter (POM) ratios from 1971-2014 (Biological C:N:P ratios project, Global POM project)
Abstract: Global Ocean POM Data Ratios from 1971-2014 (Biological C:N:P ratios project, Global POM project)
Expand/Collapse All
<ul style="list-style-type: none"> ▶ Description ▶ Acquisition Description ▶ Processing Description ▶ More Information about this dataset
<div style="border: 1px solid red; display: inline-block; padding: 2px;"> ISO RDF JSON HTML </div>

Fonte: *Biological and Chemical Oceanography Data Management Office*, 2020.

A utilização do DCAT para a padronização dos metadados pode ser verificada na Figura 25 a seguir, que apresenta parte dos metadados da Figura 24.

Figura 25 – Metadados em formato RDF/XML de registro do repositório *Biological and Chemical Oceanography Data Management Office*

```

<rdf:Description rdf:about="http://lod.bco-dmo.org/id/dataset/526747#dcatDistribution">
  <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Distribution"/>
  <dc:title>Global Ocean POM Data Ratios</dc:title>
  <dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2020-12-21T16:16:45-05:00</dc:modified>
  <dc:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2020-12-21</dc:issued>
  <dcat:downloadURL rdf:resource="https://www.bco-dmo.org/dataset/526747/data/download"/>
  <dcat:mediaType rdf:datatype="http://www.w3.org/2001/XMLSchema#string">text/tab-separated-values</dcat:mediaType>
</rdf:Description>
<rdf:Description rdf:about="https://www.bco-dmo.org/dataset/526747/data/download">
  <rdf:type rdf:resource="http://purl.org/dc/terms/FileFormat"/>
  <dc:format rdf:datatype="http://www.w3.org/2001/XMLSchema#string">text/tab-separated-values</dc:format>
  <rdfs:seeAlso rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
    https://www.bco-dmo.org/dataset/526747/data/download
  </rdfs:seeAlso>

```

Fonte: *Biological and Chemical Oceanography Data Management Office*, 2020.

No trecho da Figura 25 é possível observar a presença de dois elementos do DCAT, sendo eles o `dcat:downloadURL` e `dcat:mediaType`.

6.4 COEMS *Open Data*

O “COEMS (Continuous Observation of Embedded Multicore Systems) é uma nova plataforma de observação com métodos de verificação de suporte para sistemas embarcados multicore.” (COEMS OPEN DATA, [2020?], n.p., tradução nossa). Desse modo, o repositório COEMS *Open Data* visa o armazenamento de dados de rastreamento de *software* e *hardware*, tanto dados acadêmicos quanto industriais (COEMS OPEN DATA, [2020?]).

Assim como nos registros do repositório *Ag Data Commons*, os registros do COEMS *Open Data* apresentam a opção de visualizar os metadados em JSON e RDF/XML (Figura 26).

Figura 26 – Registro do repositório COEMS *Open Data*

The screenshot shows a dataset record with the following structure:

- License:** Open Data Commons Attribution License
- Other Access:** The information on this page (the dataset metadata) is also available in these formats.
 - JSON
 - RDF
 - via the DKAN API
- Social:** Twitter, LinkedIn, Reddit, Google+, Facebook
- Dataset Title:** Dataset AIRBUS Readers/Writers Data Race Example
- Software Trace:** This dataset shows the Software Traces produced by the execution of the instrumented AIRBUS Readers/Writers example. In addition, we provide the TeSSLa specification that we used for detecting data-race conditions in the code, and the report it generates.
- Data and Resources:**
 - Navigation: Data Race, software trace
 - Table:

Field	Value
Modified Date	2020-03-27
Release Date	2020-03-27
Identifier	13a37727-c0b2-4d45-b916-d68c8be13aa6
Temporal Coverage	Friday, March 27, 2020 - 00:00
Language	English
License	Open Data Commons Attribution License

Fonte: COEMS *Open Data*, 2020.

Ao acessar seus metadados, observa-se a utilização do DCAT, conforme a Figura 27 a seguir.

Figura 27 – Metadados em formato RDF/XML de registro do repositório COEMS *Open Data*

```

<rdf:RDF>
  <dc:Dataset rdf:about="http://dkan.isp.uni-luebeck.de/dataset/dataset-airbus-readerswriters-data-race-example">
    <dc:title>Dataset AIRBUS Readers/Writers Data Race Example</dc:title>
    <dc:description>
      <p>This dataset shows the Software Traces produced by the execution of the instrumented AIRBUS Readers/Writers example. In addition, we provide the TeSSLa specification that we used for detecting data-race conditions in the code, and the report it generates.</p> <p>The AIRBUS Readers/Writers example has three compilation flags for introducing/deactivating data race and starvation errors in the code. The "default" mode does not explicitly include any data race condition in the code, so the errors that the TeSSLa specification may detect at this stage are caused either by a false positive or an actual implementation error.</p>
    </dc:description>
    <dc:keyword>Data Race</dc:keyword>
    <dc:keyword>software trace</dc:keyword>
    <dc:theme>Software Trace</dc:theme>
    <dc:identifier>13a37727-c0b2-4d45-b916-d68c8be13aa6</dc:identifier>
    <dc:issued>2020-03-27T16:53:46+01:00</dc:issued>
    <dc:modified>2020-03-27T16:53:46+01:00</dc:modified>
    <dc:language>en</dc:language>
    <dc:distribution rdf:resource="http://dkan.isp.uni-luebeck.de/dataset/dataset-airbus-readerswriters-data-race-example/resource/06c02026-a3bf-44b8-8416"/>
    <dc:distribution rdf:resource="http://dkan.isp.uni-luebeck.de/dataset/dataset-airbus-readerswriters-data-race-example/resource/7811a3e5-4009-4508-91fa"/>
    <dc:distribution rdf:resource="http://dkan.isp.uni-luebeck.de/dataset/dataset-airbus-readerswriters-data-race-example/resource/6472fbfc-e012-424a-b59e"/>
  </dc:Dataset>

```

Fonte: COEMS *Open Data*, 2020.

Na Figura 27 acima é possível visualizar parte dos metadados do registro apresentado na Figura 26, em RDF. Nele constata-se a presença dos elementos `dc:Dataset`, `dc:keyword`, `dc:theme` e `dc:distribution`, pertencentes ao vocabulário DCAT.

6.5 Data.gov.au

O portal `data.gov.au`¹⁹ disponibiliza dados governamentais da Austrália sob acesso aberto, publicados por agências federais, estaduais e locais. Fornece, também, informações sobre dados não publicados e dados pagos, e dados de pesquisa de investigações de financiamento público que possam ser de interesse público (AUSTRALIAN GOVERNMENTDATA.GOV.AU, [2021?]).

Em seus registros encontram-se disponíveis as opções de visualizar os metadados em ISO19115/ISO19139 XML, RDF/XML e JSON, conforme ilustra a Figura 28 abaixo.

¹⁹ O catálogo do `Data.gov.au` encontra-se disponível em <https://data.gov.au/data/> e em fase *beta* em <https://data.gov.au/>. Para a análise foi considerada a versão do primeiro *link*.

Figura 28 – Registro de Data.gov.au

The screenshot shows the Data.gov.au interface for the 'State Library of Western Australia digital photographic collection'. On the left, there is a sidebar with the organization's logo and name. The main content area displays the dataset title, a brief description, and a list of resources. A red box highlights the download options: 'ISO19115/ISO19139 XML', 'RDF', and 'JSON'. The resources listed include 'SLWA Pictorial Collection' (ZIP), 'Image urls and captions' (CSV), and 'SLWA Pictorial Collection' (CSV, Marcxml, Modsxml), each with an 'Explore' button.

Fonte: Australian Government, 2016.

Na Figura 29 a seguir encontra-se parte dos metadados em RDF do registro da Figura 28.

Figura 29 – Metadados em formato RDF/XML de registro do Data.gov.au

```

<rdf:RDF>
  <dc:Dataset rdf:about="https://data.gov.au/dataset/0ea86081-2a86-43ca-87d1-e678bb55a1eb">
    <dc:publisher>
      <foaf:Organization rdf:about="https://data.gov.au/organization/c1956417-4ff6-460a-af97-4754102fd2ff">
        <foaf:name>State Library of Western Australia</foaf:name>
      </foaf:Organization>
    </dc:publisher>
    <dc:distribution>
      <dc:Distribution rdf:about="https://data.gov.au/dataset/0ea86081-2a86-43ca-87d1-e678bb55a1eb/resource/6ebc1ccf-9d1b-4ef8-b0bc-2c19543a77e9">
        <dc:format>modsxml</dc:format>
        <dc:title>SLWA Pictorial Collection (Modsxml)</dc:title>
        <dc:accessURL rdf:resource="https://data.gov.au/dataset/0ea86081-2a86-43ca-87d1-e678bb55a1eb/resource/6ebc1ccf-9d1b-4ef8-b0bc-2c19543a77e9/download/slwapictorialmods.xml"/>
      </dc:Distribution>
      <dc:description>
        This MODSXML file contains limited bibliographic data from this collection including title of image; subject; description of resource and a link to the image on the SLWA website. Tools used to create this dataset: MarcEdit MODSXML AACR2 - Anglo American Cataloguing Rules RDA - Resource Description and Access LCSH - Library of Congress Subject Headings About MODSXML: Metadata Object Description Schema (MODS) is a schema for a bibliographic element set that may be used for a variety of purposes, and particularly for library applications. The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress with input from users.
      </dc:description>
    </dc:distribution>
  </dc:Dataset>
</rdf:RDF>

```

Fonte: Australian Government, 2016.

Pode-se verificar a presença de elementos do DCAT na Figura 29 acima, sendo eles o `dc:Dataset`, `dc:distribution`, `dc:Distribution` e `dc:accessURL`.

6.6 Etsin Research Data Finder

O *Etsin Research Data Finder* é um buscador que permite encontrar dados de pesquisa e os seus metadados de diferentes áreas do conhecimento. Com a

naureza de buscador, os metadados são sempre abertos a todos, enquanto a restrição de acesso aos dados é estabelecida por seu autor (ETSIN, [2020?]).

O buscador apresenta a opção de realizar *download* dos metadados dos registros apenas em Metax JSON. No arquivo em si não é possível observar a presença do DCAT. Após contato com a plataforma, foi verificado que no modelo de dados do API Metax as associações entre os termos utilizados por ele e os termos correspondentes de outros vocabulários, incluindo o DCAT, são sinalizados e ligados (Figura 30).

Figura 30 – Elemento ‘*Research Dataset*’ do Metax

The screenshot displays the 'Research Dataset' element in Metax. At the top, there is a 'Draft' status indicator, a 'Show history' button, and an 'Export' button with a dropdown arrow. Below this, the 'Shape information' section is divided into two columns. The left column contains: 'Shape label' (Research Dataset), 'Shape id' (mrd:ResearchDataset), and 'Specialized from class' (dcat:Dataset). The right column contains: 'Description' (A collection of data, published or curated by a single source, and available for access or download in one or more formats), 'Status' (Draft), and 'Defined by profile' (Metax Research Datasets). At the bottom, a light blue box contains the text 'No concept defined'.

Fonte: *Data Vocabularies*, 2020.

Na Figura 30 acima verifica-se que o termo ‘*Research Dataset*’ do Metax corresponde a *dcat:Dataset*, elemento do DCAT. Apesar do DCAT estar presente no modelo do Metax, não é possível afirmar que o repositório utiliza, efetivamente, o vocabulário, uma vez que não o inclui diretamente em seus metadados, sendo que se trata de um dos pontos que indica a conformidade de um catálogo com o DCAT, conforme abordado na seção 5 anteriormente.

6.7 European Data Portal

O *European Data Portal* realiza a coleta de metadados de diferentes catálogos de dados da Europa, sendo seu foco informações do setor público (*Public Sector Information*). A plataforma foi fundada pela União Europeia e é mantida pelo Serviço das Publicações da União Europeia (*Publications Office of the European Union*) (EUROPEAN DATA PORTAL, 2020a).

Os registros da plataforma são disponibilizados em *Linked Data* e DQV²⁰, conforme a Figura 31 apresenta.

Figura 31 – Registro do *European Data Portal*

The screenshot shows a dataset record on the European Data Portal. At the top, there are navigation tabs: 'Dataset', 'Categories', and 'Similar Datasets'. On the right, there are buttons for 'Feedback', 'Share', and dropdown menus for 'Linked Data' and 'DQV Data'. The main title is 'Etat du trafic en temps réel' with a French flag icon and the text 'Plateforme ouverte des données publiques françaises'. The update date is 'Updated: 22.12.2020 23:59'. The description is in French, detailing real-time traffic data for Rennes Métropole, provided by Autoroutes Trafic. It includes technical details about data format (ISO 8601), location references, and various traffic status codes. At the bottom, there are two informational messages: 'There is no translation available for the desired language. The information is displayed in its default language.' and 'An updated translation of this dataset is in progress.'

Fonte: *European Data Portal*, 2020b.

Tanto para *Linked Data* quanto para DQV os formatos disponíveis são: RDF/XML, Turtle, Notation3, N-Triples e JSON-LD. A Figura 32 a seguir apresenta parte do registro da Figura 31 em JSON-LD.

²⁰ O *Data Quality Vocabulary* (DQV) é “[...] uma extensão do vocabulário DCAT para cobrir a qualidade dos dados, com que frequência são atualizados, se aceita correções do usuário, compromissos de persistência, etc.” (ALBERTONI; ISAAC, 2016, n.p., tradução nossa).

Figura 32 – Metadados em formato JSON-LD do *European Data Portal*

```

@context:
  homepage:
    @id: "http://xmlns.com/foaf/0.1/homepage"
    @type: "@id"
    label: "http://www.w3.org/2000/01/rdf-schema#Label"
    name: "http://xmlns.com/foaf/0.1/name"
    description: "http://purl.org/dc/terms/description"
    keyword: "http://www.w3.org/ns/dcat#keyword"
  distribution:
    @id: "http://www.w3.org/ns/dcat#distribution"
    @type: "@id"
    identifier: "http://purl.org/dc/terms/identifier"
    title: "http://purl.org/dc/terms/title"
  publisher:
    @id: "http://purl.org/dc/terms/publisher"
    @type: "@id"
  issued:
    @id: "http://purl.org/dc/terms/issued"
    @type: "http://www.w3.org/2001/XMLSchema#dateTime"
  modified:
    @id: "http://purl.org/dc/terms/modified"
    @type: "http://www.w3.org/2001/XMLSchema#dateTime"
    mediaType: "http://www.w3.org/ns/dcat#mediaType"
  license:
    @id: "http://purl.org/dc/terms/License"
    @type: "@id"
  downloadURL:
    @id: "http://www.w3.org/ns/dcat#downloadURL"
    @type: "@id"

```

Fonte: *European Data Portal*, 2020b.

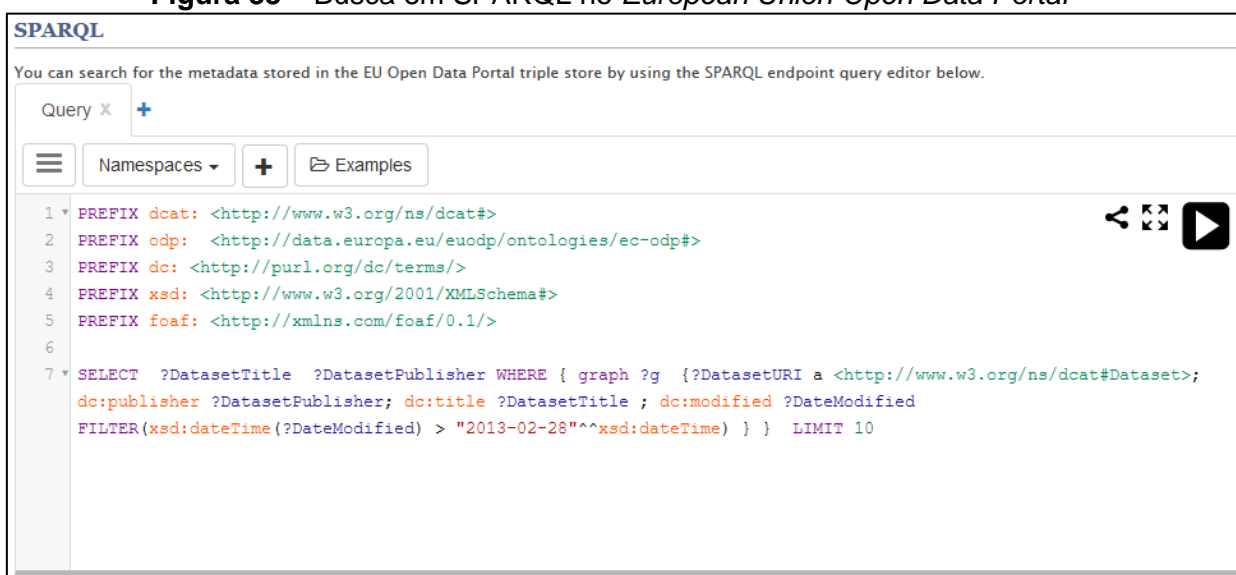
A partir da Figura 32 pode-se verificar a ligação do registro às URLs dos elementos do DCAT, como `dcat:keyword`, `dcat:distribution`, `dcat:mediaType` e `dcat:downloadURL`. Adicionalmente, o portal também permite a busca por SPARQL.

6.8 *European Union Open Data Portal*

O *European Union Open Data Portal* (EU ODP) é um catálogo de acesso aberto apresentando dados de instituições europeias e outros órgãos europeus. Os tipos de dados compreendidos pelo portal incluem: “dados geográficos, geopolíticos e financeiros; estatísticas; resultados eleitorais; atos legais; dados sobre crime, saúde, meio ambiente, transporte e pesquisa científica.” (EUROPEAN UNION OPEN DATA PORTAL, [2020?]a, n.p., tradução nossa).

O EU ODP não disponibiliza para *download* os metadados em seus registros. Entretanto, é possível realizar a busca no portal em *Linked Data*, por meio do SPARQL, conforme observa-se na Figura 33 abaixo.

Figura 33 – Busca em SPARQL no *European Union Open Data Portal*



Fonte: *European Union Open Data Portal*, [2020?]b.

Na Figura 33 encontra-se um exemplo de busca por todos os conjuntos de dados (máximo de 10) modificados após uma data específica (28 de fevereiro de 2013). Na linha 1 é possível observar o prefixo do DCAT, enquanto na linha 7 tem-se o redirecionamento a `dcat:Dataset`.

6.9 Fairdata IDA Research Data Storage Service

O *Fairdata IDA Research Data Storage Service* é organizado pelo Ministério da Educação e Cultura da Finlândia e possui como objetivo o armazenamento de dados de pesquisa. A plataforma é de acesso restrito, sendo o serviço disponibilizado de modo gratuito para “[...] universidades finlandesas e universidades de ciências aplicadas, institutos de pesquisa, bem como pesquisas financiadas pela Academia da Finlândia.” (FAIRDATA IDA, [2020?], n.p., tradução nossa). Sendo uma plataforma de acesso restrito, não foi possível realizar a sua análise.

6.10 GovData

O GovData é o portal de dados da Alemanha, tendo como objetivo fornecer acesso centralizado e padronizado a dados administrativos em nível federal, estadual e local. O portal busca a promoção do uso de licenças abertas e da disponibilização de dados legíveis por máquina. Sendo um processo em andamento, o GovData apresenta, também, dados de acesso limitado (GOVDATA, 2016).

Nos registros do portal são disponibilizados os metadados em formato RDF/XML, conforme a Figura 34.

Figura 34 – Registro do GovData

The screenshot shows a record page for 'Bevölkerung nach Altersjahren und Geschlecht' (Population by age and gender) in Bremen. The page includes a 'Datensatz' (Dataset) icon, a title, a description, and a list of metadata. A red box highlights the 'Link zu den Metadaten im Format RDF/XML' (Link to metadata in RDF/XML format) with the URL: <https://www.govdata.de/ckan/dataset/bevolkerung-nach-altersjahren-und-geschlecht.rdf>. The metadata on the right includes: 'Offenheit der Lizenz: Freie Nutzung' (Open license: Free use), 'Nutzungsbedingungen: Creative Commons Namensnennung (CC-BY)' (Usage conditions: Creative Commons Attribution), 'Letzte Änderung: 27.12.2020' (Last change: 27.12.2020), 'Veröffentlichungsdatum: 30.08.2020' (Publication date: 30.08.2020), 'Veröffentlichende Stelle: Freie und Hansestadt Bremen' (Publishing authority: Free and Hansestadt Bremen), 'Kategorien: -' (Categories: -), 'Zeitraum: 01.01.2019 - 31.12.2019' (Time period: 01.01.2019 - 31.12.2019), 'Raumbezug: -' (Spatial reference: -), and 'Schlagwörter: altersjahren, bevölkerung, bremen, geschlecht, statistik' (Keywords: ages, population, bremen, gender, statistics).

Fonte: GovData, 2020.

Ao abrir o *link* disponível observa-se a presença do DCAT, conforme a Figura 35 a seguir.

Figura 35 – Metadados em formato RDF/XML do GovData

```

<rdf:RDF>
  <dcat:Dataset rdf:about="https://www.transparenz.bremen.de/de/datensatz/bremen65.c.17154.de">
    <dcat:keyword>statistik</dcat:keyword>
    <dct:identifier>
      https://www.transparenz.bremen.de/de/datensatz/bremen65.c.17154.de
    </dct:identifier>
    <dct:temporal>
      <dct:PeriodOfTime rdf:nodeID="Nb246071d72ba4c6e8a2fe5d388427705">
        <schema:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-01-01T00:00:00</schema:startDate>
        <schema:endDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-12-31T00:00:00</schema:endDate>
      </dct:PeriodOfTime>
    </dct:temporal>
    <dct:accrualPeriodicity rdf:resource="http://publications.europa.eu/resource/authority/frequency/ANNUAL"/>
  </dcat:Distribution>
  <dcat:Distribution rdf:about="https://www.transparenz.bremen.de/de/datensatz/bremen65.c.17154.de#distribution_150185">
    <dct:license rdf:resource="http://dcat-ap.de/def/licenses/cc-by"/>
    <dct:description>
      Bevölkerung nach Geschlecht und Altersjahren (Stand 31.12.)
    </dct:description>
    <dct:language rdf:resource="http://publications.europa.eu/resource/authority/language/GER"/>
    <dct:format rdf:resource="http://publications.europa.eu/resource/authority/file-type/CSV"/>
    <dcat:accessURL rdf:resource="http://www.statistik-bremen.de/soev/abfrage_csv.cfm?tabelle=12411-04-02"/>
  </dcat:Distribution>
</rdf:RDF>

```

Fonte: GovData, 2020.

Na Figura 35 encontram-se presentes elementos do DCAT, como `dcat:Dataset`, `dcat:keyword`, `dcat:distribution`, `dcat:Distribution` e `dcat:accessURL`, indicando o uso do vocabulário pelo portal.

6.11 ICOS Carbon Portal

O *Integrated Carbon Observation System* (ICOS) é uma infraestrutura de pesquisa de gases do efeito estufa da Europa (INTEGRATED CARBON OBSERVATION SYSTEM, [2020?]a). O *ICOS Carbon Portal* realiza o armazenamento dos dados do ICOS e permite o acesso livre a esses dados, além do acesso a produtos e serviços científicos e educacionais (INTEGRATED CARBON OBSERVATION SYSTEM, [2020?]b).

Ao abrir um registro do *ICOS Carbon Portal* é possível visualizar seus metadados em HTML, JSON, RDF/XML e RDF/Turtle, conforme indica a Figura 36.

Figura 36 – Registro do *ICOS Carbon Portal*

The screenshot shows a metadata record for 'ICOS ATC NRT CO2 growing time series from Zeppelin' with the date range '2020-06-01-2020-12-27'. The record includes an 'About' section with fields like PID, Theme, Affiliation, Description, Citation, Previous version, Type, Level, File name, and Size. An 'Acquisition' section lists Station, Temporal coverage, Instrument, and Sampling height. A map of Svalbard is shown with a location pin. A 'Keywords' section lists 'CO2', 'concentration', and 'ICOS'. A 'Metadata' section at the bottom provides links for 'HTML landing page', 'JSON', 'RDF/XML', and 'RDF/TURTLE'.

About	
PID	11676/EeHjrPwgCmHTKUdswQsqX559
Theme	Atmospheric data
Affiliation	ICOS
Description	Automatically-produced, daily-growing (by data object deprecation) data on molar fraction of atmospheric CO2. The data starts from the last respective L2 release.
Citation	Lund Myhre, C., Platt, S., Hermansen, O., Lunder, C., ICOS RI, 2020. ICOS ATC NRT CO2 growing time series, Zeppelin (15.0 m), 2020-06-01-2020-12-27, https://hdl.handle.net/11676/EeHjrPwgCmHTKUdswQsqX559
Previous version	View previous version
Type	ICOS ATC NRT CO2 growing time series
Level	1
File name	ICOS_ATC_NRT_ZEP_2020-06-01_2020-12-27_15.0_591_CO2.zip
Size	66 KB

Acquisition	
Station	Zeppelin
Temporal coverage (UTC)	2020-06-01 00:00:00 - 2020-12-27 23:00:00
Instrument	ATC_591
Sampling height	15 m

Keywords
CO2 concentration ICOS

Metadata
[HTML landing page](#) • [JSON](#) • [RDF/XML](#) • [RDF/TURTLE](#)

Fonte: *ICOS Carbon Portal*, 2020.

Ao acessar os arquivos dos metadados não foi possível verificar a presença do DCAT, mas sim um vocabulário próprio do ICOS, conforme pode ser observado na Figura 37 abaixo, que apresenta os metadados em RDF/XML.

Figura 37 – Metadados em formato RDF/XML do ICOS Carbon Portal

```

<rdf:RDF>
  <rdf:Description rdf:about="https://meta.icos-cp.eu/objects/EeHJrPwgCmHTKUdswQsQX559">
    <rdf:type rdf:resource="http://meta.icos-cp.eu/ontologies/cpmeta/DataObject"/>
    <hasName>
      ICOS_ATC_NRT_ZEP_2020-06-01_2020-12-27_15.0_591_CO2.zip
    </hasName>
    <hasObjectSpec rdf:resource="http://meta.icos-cp.eu/resources/cpmeta/atcCo2NrtGrowingDataObject"/>
    <hasSha256sum rdf:datatype="http://www.w3.org/2001/XMLSchema#base64Binary">EeHJrPwgCmHTKUdswQsQX559GI3Vos0FXbF9IPNL7Tk</hasSha256sum>
    <hasSizeInBytes rdf:datatype="http://www.w3.org/2001/XMLSchema#long">67175</hasSizeInBytes>
    <wasAcquiredBy rdf:resource="http://meta.icos-cp.eu/resources/acq_EeHJrPwgCmHTKUdswQsQX559"/>
    <wasProducedBy rdf:resource="http://meta.icos-cp.eu/resources/prod_EeHJrPwgCmHTKUdswQsQX559"/>
    <wasSubmittedBy rdf:resource="http://meta.icos-cp.eu/resources/subm_EeHJrPwgCmHTKUdswQsQX559"/>
    <hasNumberOfRows rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">5040</hasNumberOfRows>
    <isNextVersionOf rdf:resource="https://meta.icos-cp.eu/objects/kWmJhJlPaKLMzdXwkWL2yY7"/>
    <hasCitationString>
      Lund Myhre, C., Platt, S., Hermansen, O., Lunder, C., ICOS RI, 2020. ICOS ATC NRT CO2 growing time series, Zeppelin (15.0 m), 2020-06-01–2020-12-27,
      https://hdl.handle.net/11676/EeHJrPwgCmHTKUdswQsQX559
    </hasCitationString>
    <hasBiblioInfo>
      {"authors":[{"firstName":"Cathrine","lastName":"Lund Myhre","orcid":"0000-0003-3587-5926","self":{"comments":[],"uri":"http://meta.icos-cp.eu/resources/people/Cathrine_Lund%20Myhre"}}, {"firstName":"Stephen Matthew","lastName":"Platt","orcid":"0000-0002-8439-6633","self":{"comments":[],"uri":"http://meta.icos-cp.eu/resources/people/Stephen%20Matthew_Platt"}}, {"firstName":"Ove","lastName":"Hermansen","orcid":"0000-0001-7353-057X","self":{"comments":[],"uri":"http://meta.icos-cp.eu/resources/people/Ove_Hermansen"}}, {"firstName":"Chris","lastName":"Lunder","self":{"comments":[],"uri":"http://meta.icos-cp.eu/resources/people/Chris_Lunder"}}],"citationString":"Lund Myhre, C., Platt, S., Hermansen, O., Lunder, C., ICOS RI, 2020. ICOS ATC NRT CO2 growing time series, Zeppelin (15.0 m), 2020-06-01–2020-12-27, https://hdl.handle.net/11676/EeHJrPwgCmHTKUdswQsQX559","temporalCoverageDisplay":"2020-06-01–2020-12-27"}
    </hasBiblioInfo>
    <rdf:Description>
  </rdf:RDF>

```

Fonte: ICOS Carbon Portal, 2020.

A plataforma também permite a busca por SPARQL, em que também não consta a presença do DCAT. Em contato com o portal, foi informado que o mapeamento de seus metadados para DCAT ainda está em progresso. Desse modo, compreende-se que, atualmente, não é possível verificar a utilização do vocabulário DCAT pelo ICOS Carbon Portal.

6.12 JRC Data Catalogue

O JRC Data Catalogue é o catálogo de dados produzidos e mantidos pelo Joint Research Centre (JRC), que por sua vez consiste no

[...] serviço científico interno da Comissão Europeia que emprega cientistas para realizar investigação, a fim de fornecer aconselhamento científico independente e apoio às políticas da União Europeia. (JOINT RESEARCH CENTRE DATA CATALOGUE, 2020, n.p., tradução nossa).

O catálogo utiliza um perfil de aplicação próprio, chamado DCAT-AP-JRC²¹, sendo que ainda não se encontra completamente implementado²².

Ao acessar os registros do JRC Data Catalogue não existe a opção de *download* ou visualização de seus metadados. Entretanto, seu acesso pode ser feito a partir da adição de '.rdf' ou '.rj' ao URL do registro, conforme informado por *e-mail* em contato realizado com o catálogo. Foi possível identificar que a extensão .rj não

²¹ Disponível em: <https://ec-jrc.github.io/dcat-ap-jrc/>.

²² Informação obtida por meio de contato por *e-mail* com a plataforma, em 02 de dezembro de 2020.

está ativa no momento. A Figura 38 a seguir ilustra parte dos metadados de um registro em RDF/XML, obtidos a partir da adição de .rdf ao seu URL.

Figura 38 – Metadados em formato RDF/XML de registro do JRC *Data Catalogue*

```
<rdf:RDF>
  <dc:Dataset rdf:about="http://data.europa.eu/89h/8141a398-41a8-42fa-81a4-5b825a51761b">
    <dc:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-10-28T10:34:07.788826</dc:issued>
  <dc:distribution>
    <dc:Distribution rdf:about="http://data.jrc.ec.europa.eu/dataset/e9963164-52d6-4ca3-a43f-6d73d9082474/resource/430d456a-8749-41a2-9759-d057209b6d7c">
      <dc:title>
        09 Baseline scenario of the total energy system up to 2050
      </dc:title>
      <dc:format>PDF</dc:format>
    <dc:description>
      JRC-EU-TIMES model outputs for the 14 MS and the EU for the project Heatroadmap 2050
    </dc:description>
    <dc:accessURL rdf:resource="https://heatroadmap.eu/wp-content/uploads/2018/11/HRE4_DS.2.pdf"/>
  </dc:Distribution>
</dc:distribution>
```

Fonte: *Joint Research Centre Data Catalogue*, 2019.


A partir da Figura 38 acima pode-se verificar a presença dos elementos `dc:Dataset`, `dc:distribution`, `dc:Distribution` e `dc:accessURL` advindos do DCAT. Observa-se que os metadados ainda não estão atualizados em conformidade com a nova versão que será lançada do DCAT-AP-JRC. Destaca-se que o fato do catálogo não disponibilizar os metadados de forma explícita desfavorece a sua acessibilidade.

6.13 Norwegian Polar Data Centre

O *Norwegian Polar Data Centre*, mantido pelo *Norwegian Polar Institute*, é responsável pela gestão e disponibilização de dados de monitoramento ambiental e de mapas topográficos e geológicos das regiões polares. O *Norwegian Polar Institute* foca em pesquisa científica e atividades de gestão ambiental no Ártico e Antártico (NORWEGIAN POLAR INSTITUTE, [2020?]; REGISTRY OF RESEARCH DATA REPOSITORIES, 2020c).


Nos registros da plataforma é possível identificar, sob '*Machine-readable data*', a opção de visualizar os metadados nos formatos JSON, DIF XML, ISO 19139 XML, Atom entry XML, JSON-LD e Datacite XML, conforme apresenta a Figura 39.

Figura 39 – Registro de Norwegian Polar Data Centre

 **Machine-readable data**
<https://doi.org/10.21334/npolar.2016.cae21585>

Formater: [JSON](#) | [DIF XML](#) | [ISO 19139 XML](#) | [Atom entry XML](#) | [DCAT \(JSON-LD\)](#) | [Datacite XML](#)

Metadata schema: <https://api.npolar.no/schema/dataset-1>

 The [Dataset catalogue](#) is hereby declared to be in the public domain. To the extent possible under Norwegian law, [Norwegian Polar Institute](#) waives all copyright and related or neighboring rights. The *Dataset catalogue* is published from Norway. See <https://doi.org/10.21334/npolar.2008.ae1a945b> for details on accessing, reusing, and harvesting the entire metadata catalogue"; Careful: this public domain dedication concerns metadata. Actual data, linked to in the metadata may have a different license.

Fonte: *Norwegian Polar Data Centre*, 2016.

A Figura 40 a seguir apresenta parte dos metadados de um registro em JSON-LD, em que se utiliza o DCAT.

Figura 40 – Metadados em formato JSON-LD de registro do Norwegian Polar Data Centre

```

@context:
  dcat: "http://www.w3.org/ns/dcat#"
  dc: "http://purl.org/dc/terms/"
  foaf: "http://xmlns.com/foaf/0.1/"
  vcard: "http://www.w3.org/2006/vcard/ns#"
  dcat:accessURL:
    @type: "@id"
  dcat:landingPage:
    @type: "@id"
  dc:license:
    @type: "@id"
  dc:publisher:
    @type: "@id"
  dc:issued:
    @type: "http://www.w3.org/2001/XMLSchema#dateTime"
  dc:modified:
    @type: "http://www.w3.org/2001/XMLSchema#dateTime"
  id: "dc:identifier"
  _rev: null
@graph:
  0:
    @id: "http://api.npolar.no/dataset/?q=&filter-id=cae21585-1e7e-4f86-b6be-a8b13696d402&format=json&variant=Ld"
    @type: "dcat:Catalog"
    dcat:dataset:
      0:
        @id: "http://api.npolar.no/dataset/cae21585-1e7e-4f86-b6be-a8b13696d402"
        _rev: "78-ff0552e629d2540e7f65f5c8b0164d2e"
        @type: "dcat:Dataset"
  
```

Fonte: *Norwegian Polar Data Centre*, 2016.

No trecho dos metadados apresentado na Figura 40 pode-se observar a ligação, dentre outros, ao vocabulário DCAT, estando presentes os elementos `dcat:Catalog` e `dcat:Dataset`.

6.14 *Open Data by Socrata*

O *Open Data by Socrata* favorece a disponibilização e o compartilhamento de dados de organizações governamentais, auxiliando em questões como tomada de decisões e operação mais eficiente (REGISTRY OF RESEARCH DATA REPOSITORIES, 2018). Apesar de estar na base do re3data, a plataforma apresenta aviso em sua página inicial de obsolência da mesma a partir de abril de 2021 (OPENDATA, [2021?]). De todo modo, seguindo com a análise dos registros do *Open Data by Socrata* foi possível constatar que não existe opção de *download* ou visualização de seus metadados. Foi feita tentativa de contato por *e-mail* com a plataforma, porém não se obteve resposta.

6.15 *Open Data Kingston*

O catálogo de dados *Open Data Kingston* busca facilitar a visualização e utilização dos dados da cidade de Kingston, Canadá, sendo que “[...] promove e fomenta princípios governamentais abertos de participação, inovação, transparência e responsabilidade.” (OPEN DATA KINGSTON, [2020?]b, n.p., tradução nossa).

Não foi possível identificar nos registros dos conjuntos de dados a opção de visualização de seus metadados. Entretanto, é possível realizar o *download* do catálogo, estando disponível em XLS, CSV, RDF/Turtle, RDF/XML e RSS *feed*, conforme destacado na Figura 41.

Figura 41 – Seção de dados do *Open Data Kingston*

The screenshot displays the Open Data Kingston interface. On the left, there is a sidebar with filters for Publisher (The Corporation of the City of Kingston, 24), Keyword (Spatial Data: 39, Census: 7, Survey: 7, Community Profiles: 6, Demographics: 6, Dissemination Areas: 6), and Theme (Urban Planning: 26, Transportation: 22, Governance: 19, Recreation: 13, Environment: 12, Shelter: 12). A red box highlights a 'Download catalog' button and a list of available formats: XLS, CSV, RDF (Turtle), RDF (XML), and RSS feed. The main content area shows four dataset cards: 'Former Municipal Boundaries', 'Airport Runway', 'Waterbody', and 'Sidewalk Surface'. Each card includes a description, license URL, and a set of tags.

Fonte: *Open Data Kingston*, [2021?]a.

Na Figura 42 a seguir encontra-se parte dos metadados do catálogo em RDF/XML.

Figura 42 – Metadados em formato RDF/XML do catálogo do *Open Data Kingston*

```

<rdf:RDF>
  <rdf:Description rdf:about="http://www.opendatasoft.com">
    <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
    <foaf:name>Opendatasoft</foaf:name>
  </rdf:Description>
  <rdf:Description rdf:about="https://opendatakingston.cityofkingston.ca/api/v2/catalog/datasets/recreation-areas-shp">
    <dct:license>
      https://www.cityofkingston.ca/documents/10180/144997/CityofKingston_OpenDataLicense.pdf
    </dct:license>
    <rdf:type rdf:resource="http://www.w3.org/ns/dcat#Distribution"/>
    <dct:description>
      shp export of https://opendatakingston.cityofkingston.ca/api/v2/catalog/datasets/recreation-areas
    </dct:description>
    <dcat:accessURL rdf:resource="https://opendatakingston.cityofkingston.ca/api/v2/catalog/datasets/recreation-areas/exports/shp"/>
    <dcat:mediaType>application/zip</dcat:mediaType>
    <dct:format>shp</dct:format>
  </rdf:Description>
  <rdf:Description rdf:nodeID="Nabc77d3f0dc94b138c8f3af802db9678">
    <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
    <rdfs:label>The Corporation of the City of Kingston</rdfs:label>
  </rdf:Description>

```

Fonte: *Open Data Kingston*, [2021?]a.

É possível observar a presença do DCAT no pequeno trecho do arquivo apresentado na Figura 42, visualizando-se os elementos `dcat:accessURL` e `dcat:mediaType`, sendo descritos os conjuntos de dados que compõem o catálogo.

6.16 Open Government Canada - Open Data

O *Open Government Canada – Open Data* visa fornecer dados governamentais do Canadá de modo aberto, de forma que as pessoas possam buscar, acessar e sugerir conjuntos de dados (OPEN GOVERNMENT CANADA, 2020).

Ao acessar os registros do portal, sob “sobre este registro” (*about this record*) é possível visualizar os seus metadados em JSON, JSON-LD e RDF/XML, conforme a Figura 43 abaixo.

Figura 43 – Registro do Open Government Canada – Open Data

Proactive Disclosure - Acts of Founded Wrongdoing

The Public Servants Disclosure Protection Act (PSDPA) offers federal public sector employees and other persons a secure and confidential process for the disclosure of serious wrongdoing in the workplace, as well as protection against reprisal.

Information to which public access is provided must describe: the wrongdoing, including the identity of the person found to have committed it if necessary in order to describe the wrongdoing adequately; the recommendations, if any, set out in any report made to the chief executive; and any corrective action taken by the chief executive or the reasons why no corrective action was taken.

Publisher - Current Organization Name: Treasury Board of Canada Secretariat

Licence: [Open Government Licence - Canada](#)

Resources

Show 10 entries

Resource Name	Resource Type	Format	Language	Links
Data Dictionary - Wrongdoing	Guide	XLS	English French	Access
Data Dictionary - Wrongdoing	Guide	JSON	English French	Access
Disclosure of Information on	Website	HTML	English	Access

Contact Email: open-ouvert@tbs-sct.gc.ca

Keywords: [Proactive Disclosure](#) | [PD](#) | [founded wrongdoing](#)

Subject: [Government and Politics](#) | [Information and Communications](#)

Maintenance and Update Frequency: As Needed

Date Published: 2017-02-14

Openness Rating: ★★★

About this Record

Record Released: 2017-03-02

Record Modified: 2021-01-04

Record ID: 6e75f19c-d19d-48aa-984e-609c8d9bc403

Metadata:

- [Link to JSON format](#)
- [DCAT \(JSON-LD\)](#)
- [DCAT \(XML\)](#)

[Atom Feed](#)

Fonte: *Open Government Canada*, 2021.

A Figura 44 a seguir apresenta parte dos metadados do registro da Figura 43 no formato RDF/XML.

Figura 44 – Metadados no formato RDF/XML do *Open Government Canada – Open Data*

```

<rdf:RDF>
  <rdf:Description rdf:about="https://open.canada.ca/dataset/6e75f19c-d19d-48aa-984e-609c8d9bc403">
    <dc:language>en</dc:language>
    <dc:format>JSON</dc:format>
    <dc:language>fr</dc:language>
    <schema:itemtype rdf:resource="http://www.w3.org/ns/dcat#Distribution"/>
    <dc:accessURL rdf:resource="https://open.canada.ca/data/recombinant-schema/wrongdoing.json"/>
    <dc:title>Data Dictionary – Wrongdoing</dc:title>
  </rdf:Description>
</dc:distribution>
<schema:itemtype rdf:resource="http://www.w3.org/ns/dcat#Dataset"/>
<dc:landingPage rdf:resource="{u'fr': u'', u'en': u''}"/>
<dc:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2017-03-02T15:40:15.544182</dc:issued>

```

Fonte: *Open Government Canada*, 2021.

Conforme pode ser visto, o *Open Government Canada – Open Data* faz uso do DCAT, estando presente logo na indicação do formato dos metadados (Figura 43). Na Figura 44 identifica-se os elementos `dcat:distribution`, `dcat:accessURL` e `dcat:landingPage`.

6.17 RDPCIDAT

O *Research Data Repository – Research Department Plasmas with Complex Interactions* (RDPCIDAT) busca tornar acessível os dados de pesquisa dos grupos de pesquisa e consórcios relacionados do Departamento de Pesquisa de Plasmas com Interações Complexas da *Ruhr Universität Bochum* (RDPCIDAT, [2021?]).

Em seus registros, os metadados encontram-se disponíveis sob “outros acessos” (*other access*) nos formatos JSON e RDF/XML, conforme a Figura 45 a seguir.

Figura 45 – Registro de RDPCIDAT

The screenshot shows a dataset page on the RDPCIDAT platform. The breadcrumb trail is 'Home / Datasets / Plasma-driven in situ production of hydrogen peroxide for biocatalysis'. The dataset title is 'Plasma-driven in situ production of hydrogen peroxide for biocatalysis' by 'Applied Microbiology' (RUB). The license is 'cc-by-40'. A red box highlights the 'Other Access' section, which states that the dataset metadata is available in JSON and RDF formats via the DKAN API. The description text is partially visible, starting with 'Peroxidases and peroxygenases are promising classes of enzymes for biocatalysis...'. Below the description are tags for 'biocatalysis', 'peroxidase', 'peroxygenase', 'plasma chemistry', and 'peroxide'. A 'Dataset Info' table is partially visible at the bottom.

Fonte: RDPCIDAT, 2020.

A Figura 46 apresenta parte dos metadados do registro da Figura 45, em RDF.

Figura 46 – Metadados no formato RDF/XML de registro do RDPCIDAT

```

<rdf:RDF>
  <dc:Dataset rdf:about="https://rdpcidat.rub.de/dataset/plasma%E2%80%90driven-%E2%80%85situ-production-hydrogen-peroxide-biocatalysis">
    <dc:title>
      Plasma-driven in situ production of hydrogen peroxide for biocatalysis
    </dc:title>
    <dc:description>
      <p>Peroxidases and peroxygenases are promising classes of enzymes for biocatalysis because of their ability to carry out one-electron oxidation reactions and stereoselective oxyfunctionalizations. However, industrial application is limited, as the major drawback is the sensitivity toward the required peroxide substrates. Herein, we report a novel biocatalysis approach to circumvent this shortcoming: in situ production of H2O2 by dielectric barrier discharge plasma. The discharge plasma can be controlled to produce hydrogen peroxide at desired rates, yielding desired concentrations. Using horseradish peroxidase, it is demonstrated that hydrogen peroxide produced by plasma treatment can drive the enzymatic oxidation of model substrates. Fungal peroxygenase is then employed to convert ethylbenzene to (R)-1-phenylethanol with an ee of >96 % using plasma-generated hydrogen peroxide. As direct treatment of the reaction solution with plasma results in reduced enzyme activity, the use of plasma-treated liquid and protection strategies are investigated to increase total turnover. Technical plasmas present a noninvasive means to drive peroxide-based biotransformations.</p>
    </dc:description>
    <dc:keyword>biocatalysis</dc:keyword>
    <dc:keyword>peroxidase</dc:keyword>
    <dc:keyword>peroxide</dc:keyword>
    <dc:keyword>peroxygenase</dc:keyword>
    <dc:keyword>plasma chemistry</dc:keyword>
    <dc:theme>Plasma in Liquids</dc:theme>
    <dc:identifier>18687f98-d046-4c8b-8e12-3f45d4b2d128</dc:identifier>
    <dc:issued>2020-11-18T11:06:06+01:00</dc:issued>
    <dc:modified>2020-11-21T19:42:24+01:00</dc:modified>
    <dc:language>en-US</dc:language>
    <dc:accessRights>potential copyright conflict with publisher</dc:accessRights>
    <dc:publisher>Applied Microbiology</dc:publisher>
    <dc:distribution rdf:resource="https://rdpcidat.rub.de/dataset/plasma%E2%80%90driven-%E2%80%85situ-production-hydrogen-peroxide-biocatalysis/resource/4ee5ca6d-f231-462b"/>
    <dc:distribution rdf:resource="https://rdpcidat.rub.de/dataset/plasma%E2%80%90driven-%E2%80%85situ-production-hydrogen-peroxide-biocatalysis/resource/7e433971-9ace-465c"/>
  </dc:Dataset>

```

Fonte: RDPCIDAT, 2020.

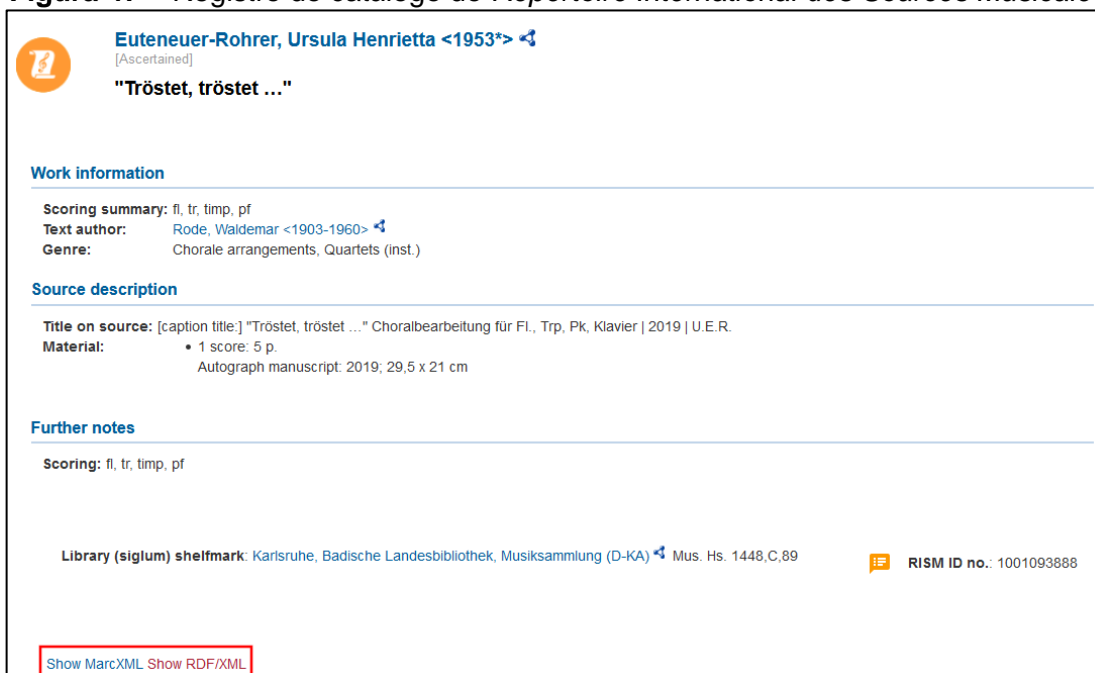
Na Figura 46 observa-se a presença de elementos do DCAT, sendo eles dcat:Dataset, dcat:keyword, dcat:theme e dcat:distribution, mostrando a utilização do vocabulário por parte do repositório.

6.18 *Répertoire International des Sources Musicales* (RISM)


O *Répertoire International des Sources Musicales* (RISM) é uma organização internacional e sem fins lucrativos que visa documentar fontes musicais, tais como impressões musicais, manuscritos musicais, libretos e teoria sobre música, podendo ser navegados a partir de seu catálogo (RÉPERTOIRE INTERNATIONAL DES SOURCES MUSICALES, [2021?]a).

Ao acessar um registro do RISM, ao fim da página apresentam-se as opções de visualizá-lo em Marc/XML e RDF/XML, conforme a Figura 47.


Figura 47 – Registro do catálogo do *Répertoire International des Sources Musicales*



The screenshot shows a catalog record for Ursula Henrietta Euteneuer-Rohrer (1953-). The work is titled "Tröstet, tröstet ...". It includes sections for work information (scoring summary, text author, genre), source description (title on source, material), and further notes (scoring). At the bottom, there are links to show the record in MarcXML and RDF/XML formats, and a library shelfmark from Karlsruhe, Badische Landesbibliothek.

Euteneuer-Rohrer, Ursula Henrietta <1953*> 
[Ascertained]
"Tröstet, tröstet ..."

Work information



Scoring summary: fl, tr, timp, pf
Text author: Rode, Waldemar <1903-1960> 
Genre: Chorale arrangements, Quartets (inst.)

Source description

Title on source: [caption title:] "Tröstet, tröstet ..." Choralbearbeitung für Fl., Trp, PK, Klavier | 2019 | U.E.R.
Material:
 • 1 score: 5 p.
 Autograph manuscript: 2019; 29,5 x 21 cm

Further notes

Scoring: fl, tr, timp, pf

Library (siglum) shelfmark: Karlsruhe, Badische Landesbibliothek, Musiksammlung (D-KA)  Mus. Hs. 1448,C,89  **RISM ID no.:** 1001093888

[Show MarcXML](#) [Show RDF/XML](#)

Fonte: *Répertoire International des Sources Musicales*, [2021?]b.

Ao acessar o registro no formato RDF/XML, não foi possível identificar a presença do DCAT, conforme a Figura 48 abaixo.

Figura 48 – Metadados no formato RDF/XML de registro do Répertoire International des Sources Musicales

```

<rdf:RDF>
  <rdf:Description rdf:about="http://data.rism.info/id/rismid/1001093888">
    <frbr:exemplar rdf:resource="http://data.rism.info/id/rismid/1001093888#ex-ks30000839-MusHs1448C89"/>
    <dc:subject>Chorale arrangements</dc:subject>
    <dc:subject>Quartets (inst.)</dc:subject>
    <dc:title>"Tröstet, tröstet ..."</dc:title>
    <dc:identifi er>1001093888</dc:identifi er>
    <cc:License rdf:resource="http://creativecommons.org/licenses/by/3.0/de"/>
    <dc:language rdf:resource="http://id.loc.gov/vocabulary/iso639-2"/>
    <dc:contributor rdf:resource="http://data.rism.info/id/rismauthorities/pe41025015"/>
    <dc:format>29,5 x 21 cm</dc:format>
    <dc:contributor>Rode, Waldemar</dc:contributor>
    <dc:extent>1 score: 5 p.</dc:extent>
    <dc:creator rdf:resource="http://data.rism.info/id/rismauthorities/pe41021775"/>
    <rda:otherTitleInformation>>false</rda:otherTitleInformation>
    <mo:arrangement_of>fl, tr, timp, pf</mo:arrangement_of>
    <dc:issued>2019</dc:issued>
    <mo:key>>false</mo:key>
  </rdf:Description>
  <rdf:Description rdf:about="http://data.rism.info/id/rismid/1001093888#ex-ks30000839-MusHs1448C89">
    <daia:label>Mus. Hs. 1448,C,89</daia:label>
    <frbr:owner rdf:resource="http://data.rism.info/id/rismauthorities/ks30000839"/>
  </rdf:Description>
</rdf:RDF>

```

Fonte: *Répertoire International des Sources Musicales*, [2021?]b.

O catálogo também oferece as opções de busca por SPARQL *endpoint* e *download* dos dados do catálogo em RDF. A Figura 49 a seguir apresenta os prefixos da busca utilizados no SPARQL *endpoint*.

Figura 49 – Busca por SPARQL *endpoint* do *Répertoire International des Sources Musicales*

```

RISM SPARQL-Endpoint

PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bsbmo:<http://bsb-muenchen.de/ont/bsbMusicOntology#>
PREFIX gnd:<http://d-nb.info/standards/elementset/gnd#>
PREFIX isbd:<http://iflastandards.info/ns/isbd/elements/>
PREFIX rdagr2:<http://metadatareregistry.org/uri/schema/RDARelationshipsGR2/>
PREFIX dc:<http://purl.org/dc/elements/1.1/>
PREFIX dcterms:<http://purl.org/dc/terms/>
PREFIX bibo:<http://purl.org/ontology/bibo/>
PREFIX mo:<http://purl.org/ontology/mo/>
PREFIX frbr:<http://purl.org/vocab/frbr/core#>
PREFIX rel:<http://purl.org/vocab/relationship/>
PREFIX rdagr1:<http://rdvocab.info/Elements/>
PREFIX umbel:<http://umbel.org/umbel#>
PREFIX daia:<http://uri.gbv.de/ontology/daia/>
PREFIX geonames:<http://www.geonames.org/ontology#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX vcard:<http://www.w3.org/2006/vcard/ns#>
PREFIX org:<http://www.w3.org/ns/org#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX cc:<https://creativecommons.org/ns#>

SELECT * WHERE {
  ?s ?p ?o
}
LIMIT 50

```

Fonte: *Répertoire International des Sources Musicales*, [2021?]c.

Os dados do catálogo que estão disponíveis para *download* em RDF são os dados de autoridade do RISM e os seus registros, bem como amostras menores de ambos²³. A Figura 50 a seguir apresenta parte dos metadados dos registros em RDF do catálogo.

²³ Disponível em: <https://opac.rism.info/main-menu-/kachelmenu/data>.

Figura 50 – Metadados no formato RDF do catálogo do *Répertoire International des Sources Musicales*

```

@prefix rdagrl: <http://rdvocab.info/Elements/> .
@prefix cc: <https://creativecommons.org/ns#> .
@prefix mo: <http://purl.org/ontology/mo/> .
@prefix geonames: <http://www.geonames.org/ontology#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix dcmitype: <http://purl.org/dc/dcmitype/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix bsbmo: <http://bsb-muenchen.de/ont/bsbMusicOntology#> .
@prefix frbr: <http://purl.org/vocab/frbr/core#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix marcrel: <http://id.loc.gov/vocabulary/relators/> .
@prefix bibo: <http://purl.org/ontology/bibo/> .
@prefix daia: <http://uri.gbv.de/ontology/daia/> .
@prefix gnd: <http://d-nb.info/standards/elementset/gnd#> .
@prefix isbd: <http://iflastandards.info/ns/isbd/elements/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

<http://data.rism.info/id/rismid/000051649>
    dct:extent "1 parts" ;
    dct:extent "5 parts" ;
    dct:extent "score: 142f." ;
    mo:key "false" ;
    dct:identifi er "000051649" ;
    frbr:exemplar <http://data.rism.info/id/rismid/000051649#ex-ks30002202-M1201619> ;
    rdagrl:otherTitleInformation "true" ;
    mo:arrangement_of "V, vl (2), vla, b, ob (2), cl (2), fag, cor (2), tr (2)" ;
    dct:format "Different sizes" ;
    cc:License <http://creativecommons.org/licenses/by/3.0/de> ;
    dct:issued "1790-1810 (18/19)" ;
    dct:description "Modern cloth binding" ;
    dct:language <http://id.loc.gov/vocabulary/iso639-2/> ;
    dct:description "Cataloguer's note, f.2v: \"xxM120.16 | Allen A Brown | 10. June, 1898\"" .

```

Fonte: *Répertoire International des Sources Musicales*, 2020.

Conforme pode ser visualizado nas Figuras 49 e 50 acima, nenhuma das duas formas de visualização e busca dos registros apresentam a utilização do DCAT. Para melhor compreender se o RISM adota o vocabulário, foi realizado contato por *e-mail*, sendo que não houve retorno conclusivo. Dessa forma, não é possível afirmar que o catálogo do RISM utiliza o DCAT.

6.19 Repositorio de datos del Ministerio de Educación del Perú

O *Repositorio de datos del Ministerio de Educación del Perú* busca fornecer acesso gratuito aos dados produzidos pelas dependências do Ministério da Educação do Peru, pensando-se na utilização, reutilização e compartilhamento de dados governamentais (MINISTERIO DE EDUCACIÓN, [2021?]).

Ao acessar um registro do repositório identifica-se a mensagem: “Esses campos são compatíveis com DCAT, um vocabulário RDF projetado para facilitar a interoperabilidade entre catálogos de dados publicados na web.” (MINISTERIO DE

EDUCACIÓN, 2019, n.p., tradução nossa). Entretanto, o repositório não disponibiliza os metadados para *download*, sendo que a plataforma não respondeu ao contato realizado por *e-mail*.

Todavia, foi possível acessar o código-fonte do registro em formato XHTML+RDFa por meio de função do próprio navegador, resultando na visualização apresentada na Figura 51 abaixo.

Figura 51 – Metadados em formato XHTML+RDFa de registro do *Repositorio de datos del Ministerio de Educación del Perú*

```

175 <div class="content">
176 <div id="data-and-resources"><h2>Data and Resources</h2><div property="dcat:distribution"><div class="item-list"><ul
class="resource-list"><li class="first"><div property="dcat:Distribution"><a href="/dataset/indicadores-de-educaci%C3%B3n-
b%C3%A1sica-alternativa/resource/9e8a80e5-85f6-4eed-babd-90f33a53db68" class="heading" title="Diccionario de datos de Indicadores
sobre Docentes y Estudiantes de EBA" property="dcat:accessURL">Diccionario de datos de Indicadores sobre Docentes y Estudiantes de
EBA<span class="format-label" property="dc:format" data-format="pdf">pdf</span></a><p class="description"></p><span
class="links"><a href="/dataset/indicadores-de-educaci%C3%B3n-b%C3%A1sica-alternativa/resource/9e8a80e5-85f6-4eed-
babd-90f33a53db68" class="btn btn-primary">Explore Data</a><a href="http://datos.minedu.gob.pe/sites/default/files/
Diccionario%20de%20datos%20de%20facilitadores%20y%20estudiantes%20EBA.pdf" class="btn btn-primary"><i class="fa fa-download">
<span> Download</span></i></a></span></div></li>
177 <li class="last"><div property="dcat:Distribution"><a href="/dataset/indicadores-de-educaci%C3%B3n-b%C3%A1sica-alternativa
/resource/6c3f381a-92c9-4d0c-b70b-40075423ca53" class="heading" title="Indicadores sobre docentes y estudiantes de EBA"
property="dcat:accessURL">Indicadores sobre docentes y estudiantes de EBA<span class="format-label" property="dc:format" data-
format="csv">csv</span></a><p class="description"></p><span class="links"><a href="/dataset/indicadores-de-educaci%C3%B3n-
b%C3%A1sica-alternativa/resource/6c3f381a-92c9-4d0c-b70b-40075423ca53" class="btn btn-primary">Explore Data</a><a
href="http://datos.minedu.gob.pe/sites/default/files/Indicadores%20sobre%20docentes%20y%20estudiantes%20EBA_0.csv" class="btn btn-
primary"><i class="fa fa-download"> <span> Download</span></i></a></span></div></li>
178 </ul></div></div></div><div class="field field-name-field-tags field-type-taxonomy-term-reference field-label-hidden"><div
class="field-items"><div class="field-item even" property="dcat:theme"><a href="/tags/educaci%C3%B3n-b%C3%A1sica-alternativa"
typeof="skos:Concept" property="rdfs:label skos:prefLabel" datatype="">EDUCACIÓN BÁSICA ALTERNATIVA</a></div><div class="field-
item odd" property="dcat:theme"><a href="/tags/instituciones-educativas" typeof="skos:Concept" property="rdfs:label
skos:prefLabel" datatype="">INSTITUCIONES EDUCATIVAS</a></div><div class="field-item even" property="dcat:theme"><a href="/tags/
educaci%C3%B3n-permanente" typeof="skos:Concept" property="rdfs:label skos:prefLabel" datatype="">EDUCACIÓN PERMANENTE</a>
</div><div class="field-item odd" property="dcat:theme"><a href="/tags/educaci%C3%B3n-pt%C3%BAblica" typeof="skos:Concept"
property="rdfs:label skos:prefLabel" datatype="">EDUCACIÓN PÚBLICA</a></div><div class="field-item even" property="dcat:theme"><a
href="/tags/educaci%C3%B3n-privada" typeof="skos:Concept" property="rdfs:label skos:prefLabel" datatype="">EDUCACIÓN PRIVADA</a>
</div></div></div><section class="field-group-table group additional"><h2>Dataset Info</h2><span class="field-group-table-
description"><i class="icon-large icon-info-sign"></i>These fields are compatible with <a href="http://www.w3.org/TR/vocab-
dcat/">DCAT</a>, an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.</span>
<table class="field-group-format group additional"><thead><tr><th>Field</th><th>Value</th></tr></thead><tr><th class="field-
label">Publisher</th><td class="field-items"><div class="field field-name-og-group-ref field-type-entityreference field-label-
hidden"><div class="field-items"><div class="field-item even" property="dc:publisher"><a href="/group/educaci%C3%B3n-b%C3%A1sica-
alternativa">Educación Básica Alternativa</a></div></div></tr><th class="field-label">Modified Date</th><td
class="field-items"><div property="dcterms:modified" class="field-name-field-modified-date">2019-11-19</div></td></tr><tr><th
class="field-label">Release Date</th><td class="field-items"><div property="dcterms:issued" class="field-name-field-release-
date">2019-11-19</div></td></tr><tr><th class="field-label">Identifier</th><td class="field-items"><div
property="dcterms:identifier" class="field-name-field-identifier">eb938eb0-3eb0-47c6-a398-bfdac2d44f</div></td></tr><tr><th
class="field-label">Temporal Coverage</th><td class="field-items"><div class="field field-name-field-temporal-coverage field-type-
datetime field-label-hidden"><div class="field-items"><div class="field-item even" property="dc:temporal"><span class="date-
display-single" property="dc:temporal" datatype="xsd:dateTime" content="2019-11-19T10:00:00+00:00">Martes, Noviembre 19, 2019 -
10:00</span></div></div></td></tr><tr><th class="field-label">License</th><td class="field-items"><div class="field field-
name-field-license field-type-text field-label-hidden"><div class="field-items"><div class="field-item even"
property="dcl:license">License Not Specified</div></div></td></tr><tr><th class="field-label">Author</th><td class="field-
items"><div class="field field-name-field-author field-type-text field-label-hidden"><div class="field-items"><div class="field-
item even" property="dc:creator">MINEDU</div></div></td></tr></table></section><div class="field field-name-field-rating
field-type-fivestar field-label-inline clearfix"><div class="field-label">Rating:<span></div><div class="field-items"><div
class="field-item even"><div class="clearfix fivestar-average-stars fivestar-average-text"><div class="fivestar-static-item"><div
class="form-item form-type-item form-group">

```

Fonte: *Ministerio de Educación*, 2019.

A Figura 51 mostra parte dos metadados de um registro do repositório em XHTML+RDFa, que faz parte da família de linguagens de marcação *eXtensible Hypertext Markup Language* (XHTML) e faz uso do *Resource Description Framework in Attributes Core* (RDFa) para a definição de atributos e sintaxe (MCCARRON, 2015). Basicamente, o XHTML+RDFa objetiva facilitar a apresentação de conteúdo, enquanto promove descrições semânticas (MCCARRON, 2015; TOMOYOSE, 2018).

Apesar do repositório não disponibilizar os metadados, foi possível verificar a utilização do DCAT a partir da presença dos elementos `dcat:distribution`, `dcat:accessURL`, `dcat:Distribution` e `dcat:theme` em XHTML+RDFa.

Assim como no caso do *JRC Data Catalogue*, a ausência do acesso direto aos metadados do repositório diminui a recuperação da informação.

6.20 Discussão sobre as análises das plataformas

A partir do levantamento realizado no re3data, em busca de plataformas que afirmassem adotar o DCAT, obteve-se um resultado de 19 portais, cerca de 0,73% do número total de plataformas registradas, demonstrando baixa aderência ao vocabulário. Por meio da análise de cada uma dessas 19 plataformas, verificou-se que quatro não utilizam o DCAT no momento dessa investigação, sendo elas o *Etsin Research Data Finder*, o *ICOS Carbon Portal*, o *Open Data by Socrata* e o *Répertoire International des Sources Musicales (RISM)*, enquanto uma é de acesso restrito, não sendo possível verificar os seus registros, sendo o caso do *Fairdata IDA Research Data Storage Service*.

Observa-se que uma parte das plataformas são voltadas a dados governamentais (*European Data Portal*, *GovData*, *Open Data by Socrata*, *Open Data Kingston*, *Open Government Canada - Open Data* e *Repositorio de datos del Ministerio de Educación del Perú*), enquanto uma parte possui como foco os dados de pesquisa (*Arctic Permafrost Geospatial Centre*, *Biological and Chemical Oceanography Data Management Office*, *COEMS Open Data*, *Etsin Research Data Finder*, *ICOS Carbon Portal*, *JRC Data Catalogue*, *Norwegian Polar Data Centre*, *RDPCIDAT* e *RISM*), conforme explorado nas subseções anteriores. Destaca-se, ainda, que algumas das plataformas disponibilizam ambos os tipos de dados (*Ag Data Commons*, *Data.gov.au* e *European Union Open Data Portal*).

No Quadro 37 apresentam-se os resultados obtidos a partir da análise, retomando-se que foram consideradas como plataformas que adotam o DCAT aquelas que utilizam em seus metadados, disponibilizados em formatos compatíveis com o DCAT, os termos do vocabulário. Na primeira coluna estão listadas as plataformas, em ordem alfabética; na segunda coluna está indicado se ela faz ou não a adoção do DCAT; e na terceira coluna é indicada a forma como foi a averiguação final de uso dele, sendo pelo próprio portal ou por *e-mail*.

Quadro 37 – Resultado da análise de utilização do *Data Catalog Vocabulary (DCAT)* nas plataformas registradas no re3data

Plataforma	Uso do DCAT	Verificação
Ag Data Commons	Sim	Portal

Arctic Permafrost Geospatial Centre	Sim	Portal
Biological and Chemical Oceanography Data Management Office	Sim	Portal
COEMS Open Data	Sim	Portal
Data.gov.au	Sim	Portal
Etsin Research Data Finder	Não*	<i>E-mail</i>
European Data Portal	Sim	Portal
European Union Open Data Portal	Sim	Portal
Fairdata IDA Research Data Storage Service	-	-
GovData	Sim	Portal
ICOS Carbon Portal	Não	<i>E-mail</i>
JRC Data Catalogue	Sim	<i>E-mail</i>
Norwegian Polar Data Centre	Sim	Portal
Open Data by Socrata	Não	Portal
Open Data Kingston	Sim	Portal
Open Government Canada - Open Data	Sim	Portal
RDPCIDAT	Sim	Portal
Répertoire International des Sources Musicales (RISM)	Não	<i>E-mail</i>
Repositorio de datos del Ministerio de Educación del Perú	Sim	Portal

Fonte: elaboração da autora.

No caso do *Etsin Research Data Finder*, a adoção do DCAT ocorre a partir do API Metax, não sendo possível verificar a utilização nos registros da plataforma em si, diferenciando-se dos demais casos que fazem o uso do vocabulário. Desse modo, optou-se por sinalizá-lo com um asterisco no Quadro 37, indicando a adoção indireta do vocabulário.

Outro caso a ser observado é o do *Ag Data Commons*, que apesar de utilizar o DCAT e ser possível verificar a sua presença em formatos JSON e RDF/XML, os registros analisados apresentam erros de validação, fator que impede que benefícios, tais como a efetiva ligação de dados, sejam atingidos.

Retoma-se que a tentativa de contato por *e-mail* foi feita com seis plataformas, sendo elas: *Etsin Research Data Finder*, *ICOS Carbon Portal*, *JRC Data Catalogue*, *Open Data by Socrata*, *Répertoire International des Sources Musicales (RISM)* e *Repositorio de datos del Ministerio de Educación del Perú*. Com exceção do *Open Data by Socrata* e do *Repositorio de datos del Ministerio de Educación del Perú*, todas as plataformas anteriores deram uma devolutiva quanto à utilização do DCAT, permitindo chegar às conclusões de uso feitas em suas respectivas subseções.

7 CONSIDERAÇÕES FINAIS

A representação da informação é central quando se aborda a disponibilização e o compartilhamento de quaisquer fontes e recursos informacionais. A partir da utilização de metadados permitem-se a identificação, preservação, acesso e recuperação da informação, sendo indispensável a adoção de vocabulários para a padronização desses metadados para a efetivação de seus benefícios.

A presente pesquisa teve como objetivo explorar a utilização dos vocabulários no contexto dos dados de pesquisa, sendo motivada pela presença cada vez maior desses como fontes de informação, tornando importante discutir a forma apropriada de tratar dos mesmos para o seu melhor aproveitamento.

Além disso, alinhados aos princípios *Linked Data*, os vocabulários veiculam a ligação dos recursos aumentando a encontrabilidade de novas informações. Um ponto que pode ser observado é que a etapa de classificação e atribuição de metadados aos dados está presente tanto nos ciclos de vida dos dados de pesquisa quanto no ciclo de vida *Linked Data*, sinalizando a convergência da importância da padronização em ambos os casos, sendo que no ciclo de vida *Linked Data* ressalta-se a utilização de ontologias.

Considera-se que foi possível atingir os objetivos delineados, tendo sido realizado o 1) aprofundamento teórico sobre a temática de dados de pesquisa, *Linked Data* e vocabulários, com a 2) investigação do DCAT, DC Terms, FOAF Vocabulary, VOID, *vCard Ontology* e PROV-O, bem como do RDF Schema, XML Schema e SKOS. Além disso, no decorrer da dissertação foi explorada a publicação de dados no contexto dos princípios *Linked Data*. A 3) análise da utilização do DCAT foi feita a partir do levantamento das plataformas registradas no re3data que afirmam adotá-lo como vocabulário. Com isso, atingiu-se o objetivo geral da pesquisa de analisar as características e os fundamentos do DCAT para a publicação de dados de pesquisa nos princípios *Linked Data*.

Verifica-se que até o momento as discussões quanto ao DCAT são escassas quando comparado aos demais vocabulários, principalmente em âmbito nacional, sendo que isso pode ser justificado pela sua recente data de publicação, em 2014, enquanto vocabulários como o *Dublin Core* possuem uma existência maior.

A partir do levantamento realizado no re3data foi possível identificar 19 resultados que afirmam adotar o DCAT, sendo que dessas, 18 são as que puderam ser analisadas, enquanto 14 são as plataformas que de fato utilizam o vocabulário e

uma utiliza indiretamente. Apesar de em pequeno número, a adoção do DCAT por essas plataformas indica a preocupação com questões de encontrabilidade, acessibilidade, interoperabilidade e reutilização, uma vez que utilizam o vocabulário para a disponibilização aberta dos dados e em formatos que favorecem os princípios FAIR, indo de encontro, ainda, com os princípios *Linked Data*.

Observa-se que o fato de nem todas as plataformas disponibilizarem os registros para *download* ou busca, como o SPARQL, dificultou o processo de análise, sendo que em alguns casos foi necessário realizar contato com elas para compreender se a plataforma adotava ou não o DCAT como vocabulário. Nota-se que nem todas as plataformas deram uma devolutiva, não sendo possível, nesses casos, completar a análise. Em um dos casos (*Fairdata IDA Research Data Storage Service*), a plataforma era de acesso restrito, não sendo possível prosseguir com a análise.

A partir da exploração e discussões realizadas, pode-se afirmar que a utilização do DCAT como vocabulário para a publicação de dados de pesquisa diminui problemas como ambiguidade e auxilia no processo de descoberta de dados e serviços de dados, uma vez que busca a padronização dos termos de descrição dos catálogos, conjuntos de dados e dos relacionamentos que podem ser estabelecidos. Sendo desenvolvido para o contexto dos catálogos e seus dados, o DCAT se apresenta um conjunto de termos conveniente e abrangente para descrever conjuntos de dados no geral, trazendo elementos como *dcat:dataset*, *dcat:service*, *dcat:catalog*, dentre outros, diferenciando-o de outros vocabulários desenvolvidos para contextos não específicos aos conjuntos de dados e catálogos. Para campos do conhecimento que necessitam de maiores detalhamentos podem ser desenvolvidos e utilizados os perfis de aplicação.

Ao adotar o DCAT em conjunto a tecnologias e conceitos que norteiam a ligação de dados, tais como a publicação dos conjuntos de dados de pesquisa em RDF e a disponibilização aberta dos dados e metadados, permite-se que sejam melhor aproveitados os benefícios dos dados de pesquisa, uma vez que amplia o acesso, a descoberta, a recuperação e a (re)utilização dos conjuntos de dados.

Tendo em vista que a Ciência da Informação opera intimamente com a representação da informação, a presente pesquisa buscou discuti-la no contexto dos dados de pesquisa, de modo a acompanhar a crescente demanda de compartilhamento e (re)utilização deles. Além disso, apresentou-se a discussão de

aspectos e ferramentas semânticas na *Web*, tais como os princípios *Linked Data*, que também são de grande relevância para a área, especialmente quando abordada a representação da informação, visto que se interpõem com os aspectos de melhoria do alcance, da visibilidade, da utilização, do acesso e da recuperação da informação. Desse modo, a pesquisa propicia e impulsiona a inovação e a visibilidade da temática no contexto da Ciência da Informação, em especial, no âmbito nacional.

Nesse sentido, o profissional da informação deve se ocupar em atualizar seus conhecimentos para estar apto a lidar com o tratamento dos dados de pesquisa, com os pesquisadores que desejam disponibilizar e compartilhar os seus dados e com os usuários em suas demandas de busca e acesso deles, impactando positivamente a sociedade.

Sugere-se para futuras pesquisas o estudo mais aprofundado do DCAT, como a comparação de utilização do DCAT frente a outros vocabulários, além da exploração de sua aplicabilidade em âmbito nacional para a publicação de dados de pesquisa. Outra sugestão é a discussão do DCAT no contexto dos perfis de aplicação, uma vez que o vocabulário tem a característica de reaproveitamento de termos de vocabulários pré-existentes, que também é uma característica dos perfis de aplicação.

REFERÊNCIAS

- AKAR, Z. *et al.* Querying the Web of interlinked datasets using Void descriptions. *In: LINKED DATA ON THE WEB, 2012, Lyon. Proceedings [...].* Lyon: [s.n.], 2012. Disponível em: <http://events.linkeddata.org/ldow2012/papers/ldow2012-paper-06.pdf>. Acesso em: 28 jul. 2020.
- ALAM, A. W. Dublin Core metadata for research data: lessons learned in a real-world scenario with datorium. *In: DCMI INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2014, Austin. Proceedings [...].* Austin, TX: DCMI, 2014. p. 64-73. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3703/1926>. Acesso em: 19 jun. 2020.
- ALBERTONI, R.; ISAAC, A. **Data on the Web Best Practices: Data Quality Vocabulary.** 2016. Disponível em: <https://www.w3.org/TR/2016/NOTE-vocab-dqv-20161215/>. Acesso em: 23 dez. 2020.
- ALBERTONI, R. *et al.* (ed.). **Data Catalog Vocabulary (DCAT) - version 2.** 2020. Disponível em: <https://www.w3.org/TR/2020/REC-vocab-dcat-2-20200204/>. Acesso em: 07 maio 2020.
- ALLEMANG, D.; HENDLER, J. **Semantic Web for the working ontologist: effective modeling in RDFS and OWL.** Amsterdam: Morgan Kaufmann, 2008.
- ALVES, R. C. V. **Metadados como elementos do processo de catalogação.** 2010. 134f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília, 2010. Disponível em: <http://hdl.handle.net/11449/103361>. Acesso em: 06 out. 2020.
- ALVES, R. C. V. **Web semântica: uma análise focada no uso de metadados.** 2005. 180 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília, 2005. Disponível em: <http://hdl.handle.net/11449/93690>. Acesso em: 13 jan. 2020.
- ARAKAKI, F. A. **Linked Data: ligação de dados bibliográficos.** 2016. 144 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Julio de Mesquita Filho, Marília, 2016. Disponível em: <http://hdl.handle.net/11449/147979>. Acesso em: 28 ago. 2019.
- ARAKAKI, F. A. **Metadados administrativos e a proveniência dos dados: modelo baseado na família PROV.** 2019. 140 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista Julio de Mesquita Filho, Marília, 2019. Disponível em: <http://hdl.handle.net/11449/180490>. Acesso em: 07 jan. 2020.
- ARAÚJO, D. G. *et al.* Contribuições para a gestão de dados científicos: análise comparativa entre modelos de ciclo de vida dos dados. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 32-51, nov. 2019. DOI: <https://doi.org/10.18617/liinc.v15i2.4686>. Disponível em: <http://revista.ibict.br/liinc/article/view/4686/4307>. Acesso em: 14 out. 2020.

ARCTIC PERMAFROST GEOSPATIAL CENTRE. **About**. [2020?]. Disponível em: <https://apgc.awi.de/about>. Acesso em: 22 dez. 2020.

ARCTIC PERMAFROST GEOSPATIAL CENTRE. **[Registro do dataset 'Historic Lake Drainage from Remote Sensing Datasets, 1955-2017, Western Arctic Coastal Plain, Northern Alaska (US)']**. 2019. Disponível em: <https://apgc.awi.de/dataset/historic-ld-rs-1955-2017-northern-alaska>. Acesso em: 22 dez. 2020.

AUER, S. *et al.* Managing the life-cycle of Linked Data with the LOD2 stack. *In*: Cudré-Mauroux P. *et al.* (Ed.). THE SEMANTIC WEB – ISWC 2012, 11., 2012, Boston. **Proceedings** [...]. Berlin Heidelberg: Springer, 2012. Disponível em: https://link.springer.com/chapter/10.1007/978-3-642-35173-0_1. Acesso em: 28 abr. 2020.

AUSTRALIAN GOVERNMENT. **About**. [2021?]. Disponível em: <https://data.gov.au/page/about>. Acesso em: 04 jan. 2021.

AUSTRALIAN GOVERNMENT. **[Registro do dataset 'State Library of Western Australia digital photographic collection']**. 2016. Disponível em: <https://data.gov.au/data/dataset/slwa-digital-photographic-collection>. Acesso em: 04 jan. 2021.

BAKER, T.; VANDENBUSSCHE, P.; VATANT, B. Requirements for vocabulary preservation and governance. **Library Hi Tech**, v. 31, n. 4, p. 657-668, 2013.

BAKER, T. *et al.* **Library Linked Data Incubator Group final report**. 2011. Disponível em: <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>. Acesso em: 05 out. 2018.

BELTRAN, A. G. *et al.* (ed.). **Data Catalog Vocabulary (DCAT) - revised edition**. 2018. Disponível em: <https://www.w3.org/TR/2018/WD-vocab-dcat-2-20180508/>. Acesso em: 09 ago. 2018.

BERNERS-LEE, T. **Content negotiation of content-type**. 2009. Disponível em: <https://www.w3.org/DesignIssues/Conneg>. Acesso em: 26 mar. 2020.

BERNERS-LEE, T. **Linked Data**. 2006. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 3 jul. 2018.

BERNERS-LEE, T. **Semantic Web**. 2000. 17 slides. XML-2000 Conference. Disponível em: <https://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html>. Acesso em: 21 out. 2020.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific american**, v. 284, n. 5, p. 34-43, 2001.

BIOLOGICAL AND CHEMICAL OCEANOGRAPHY DATA MANAGEMENT OFFICE. **About us**. c2020. Disponível em: <https://www.bco-dmo.org/about-us>. Acesso em: 22 dez. 2020.

BIOLOGICAL AND CHEMICAL OCEANOGRAPHY DATA MANAGEMENT OFFICE. **[Registro do dataset 'Global Ocean POM Data Ratios']**. 2020. Disponível em: <https://www.bco-dmo.org/dataset/526747>. Acesso em: 22 dez. 2020.

BIZER, C. The emerging Web of Linked Data. **IEEE Intelligent Systems**, [S.l.], v. 24, n. 5, set./out. 2009. DOI: <https://10.1109/MIS.2009.102>.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. **Linked Data: the story so far**. 2009. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 10 ago. 2018.

BRASIL. **Conheça as instituições de fomento à pesquisa no País**. 2017. Disponível em: <http://www.brasil.gov.br/editoria/educacao-e-ciencia/2010/09/conheca-as-instituicoes-de-fomento-a-pesquisa-no-pais>. Acesso em: 7 ago. 2018.

BRICKLEY, D.; GUHA, R. V. (ed.). **RDF Schema 1.1**. 2014. Disponível em: <http://www.w3.org/TR/2014/REC-rdf-schema-20140225/>. Acesso em: 20 mar. 2020.

BORKO, H. Information science: what is it? **American Documentation**, Washington, v. 19, n. 1, p. 3-5, jan. 1968. DOI: <https://doi.org/10.1002/asi.5090190103>.

BOURDIEU, P. Le champ scientifique. **Actes de la Recherche en Sciences Sociales**, n. 2/3, jun. 1976, p. 88-104. Tradução de Paula Montero. Disponível em: <https://cienciaetecnosociedade.files.wordpress.com/2015/05/o-campo-cientifico-pierre-bourdieu.pdf>. Acesso em: 02 jul. 2020.

BRICKLEY, D.; MILLER, L. **FOAF vocabulary specification 0.99**. 2014. Disponível em: <http://xmlns.com/foaf/spec/>. Acesso em: 27 jun. 2018.

BUCKLAND, M.K. Information as thing. Tradução livre de Luciane Artêncio. **Journal of the American Society for Information Science**, v. 45, n. 5, p. 351-360, 1991. Disponível em: https://edisciplinas.usp.br/pluginfile.php/2532319/mod_resource/content/1/Informa%C3%A7%C3%A3oComoCoisa.pdf?fbclid=IwAR02PT0A7vXi4Q_cvo1syBCHkI0IM3JzXFjLe4nL14v1xpqhSjFHvLPlaoE. Acesso em: 19 abr. 2019.

BUSH, V. As we may think. **Atlantic Monthly**, v. 176, n. 1, p. 101-108, 1945. Disponível em: <https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>. Acesso em: 19 abr. 2019.

CAPURRO, R. Epistemologia y Ciencia de la Informacion. Tradução de Ana Maria Rezende Cabral *et al.* In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., 2003, Belo Horizonte. **Anais [...]**. Belo Horizonte: [s.n.], 2003. Disponível em: http://www.capurro.de/enancib_p.htm. Acesso em: 06 ago. 2020.

CAPURRO, R.; HJØRLAND, B. The concept of information. **Annual Review of Information Science and Technology**, v. 37, chapter 8, p. 343-411, 2003. DOI: <https://doi.org/10.1002/aris.1440370109>.

CARVALHO, E. R. S.; LEITE, F. C. L. Produção/obtenção, compartilhamento e (re)uso de dados de pesquisa. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. **Anais [...]**. Florianópolis: UFSC, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/478/917>. Acesso em: 04 jun. 2020.

CASTRO, F. F. **Padrões de representação e descrição de recursos informacionais em bibliotecas digitais na perspectiva da ciência da informação**: uma abordagem do MarcOnt initiative na era da web semântica. 2008. 201 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília, 2008. Disponível em: <http://hdl.handle.net/11449/93689>. Acesso em: 16 jan. 2020.

CASTRO, F. F.; SIMIONATO, A. C.; ZAFALON, Z. R. Aspectos relacionais entre ontologia e metadados: considerações interdisciplinares. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Bahia. **Anais [...]**. Salvador: PPGCI/UFBA, 2016. p. 3916-3939. Disponível em: <http://www.ufpb.br/evento/iti/ocs/index.php/enancib2016/enancib2016>. Acesso em: 02 abr. 2018.

CERRAO, N. G. **Análise das aplicações de metadados baseada em FRBR e RDA em repositórios institucionais digitais**: contribuições do domínio bibliográfico. 2019. 122 f. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de São Carlos, São Carlos, 2019. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/11920>. Acesso em: 27 jul. 2020.

CHENEY, J.; MISSIER, P.; MOREAU, L. (ed.). **Constraints of the PROV Data Model**. 2013. Disponível em: <https://www.w3.org/TR/2013/REC-prov-constraints-20130430/>. Acesso em: 15 jul. 2020.

COEMS OPEN DATA. **[Registro do dataset 'Dataset AIRBUS Readers/Writers Data Race Example']**. 2020. Disponível em: <http://dkan.isp.uni-luebeck.de/dataset/dataset-airbus-readerswriters-data-race-example>. Acesso em: 22 dez. 2020.

COEMS OPEN DATA. **Welcome to COEMS Open Data Portal**. [2020?]. Disponível em: <http://dkan.isp.uni-luebeck.de/>. Acesso em: 22 dez. 2020.

CORDEIRO, D. *et al.* Da Ciência à eCiência: paradigmas da descoberta do conhecimento. **Revista USP**, São Paulo, v. 97, p. 71-81, mar./maio 2013. DOI: <https://doi.org/10.11606/issn.2316-9036.v0i97p71-81>. Disponível em: <http://www.revistas.usp.br/revusp/article/view/61867/64710>. Acesso em: 02 jul. 2020.

CÓRDULA, F. R.; ARAÚJO, W. J. O compartilhamento de dados científicos na era do e-science. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados científicos**:

perspectivas e desafios. João Pessoa: Ed. UFPB, 2019. p. 189-207. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/view/359/508/2949-1>. Acesso em: 29 jun. 2020.

COSTELLO, M. J. Motivating online publication of data. **BioScience**, v. 59, n. 5, p. 418-427, maio 2009. DOI: <https://doi.org/10.1525/bio.2009.59.5.9>. Disponível em: <https://academic.oup.com/bioscience/article/59/5/418/297578>. Acesso em: 10 ago. 2018.

COYLE, K. **FRBR, before and after**: a look at our bibliographic models. Chicago: ALA editions, 2016.

COYLE, K.; BAKER, T. **Guidelines for Dublin Core Application Profiles**. 2009. Disponível em: <http://www.dublincore.org/specifications/dublin-core/profile-guidelines/>. Acesso em: 30 mar. 2020.

CURTY, R. G. Abordagens de reuso e a questão da reusabilidade dos dados científicos. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 177-193, nov. 2019. DOI: <https://doi.org/10.18617/liinc.v15i2.4777>. Disponível em: <http://revista.ibict.br/liinc/article/view/4777/4315>. Acesso em: 14 out. 2020.

CURTY, R. G.; AVENTURIER, P. O paradigma da publicação de dados e suas diferentes abordagens. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. **Anais [...]**. Marília: Unesp, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/105144>. Acesso em: 14 out. 2020.

CURTY, R. G. *et al.* Attitudes and norms affecting scientists' data reuse. **PLoS ONE**, Bloomington-IN, v. 12, n. 12, p. 1-22, dez. 2017. DOI: <https://doi.org/10.1371/journal.pone.0189288>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189288>. Acesso em: 14 out. 2020.

CYGANIAK *et al.* **Vocabulary of Interlinked Datasets (VOID)**. 2011. Disponível em: <http://vocab.deri.ie/void>. Acesso em: 14 abr. 2020.

DATA VOCABULARIES. **Metax research datasets**: research dataset. 2020. Disponível em: <https://tietomallit.suomi.fi/model/mrd/ResearchDataset/>. Acesso em: 23 dez. 2020.

DATA CITE. **Mission**. [2020?]. Disponível em: <https://datacite.org/mission.html>. Acesso em: 07 jan. 2020.

DBPEDIA. **About**. c2019. Disponível em: <https://wiki.dbpedia.org/about>. Acesso em: 15 abr. 2020.

DCMI USAGE BOARD. **DCMI Metadata Terms**. 2020. Disponível em: <http://dublincore.org/documents/dcmi-terms/>. Acesso em: 13 mar. 2020.

DIAS, G. A.; ANJOS, R. L.; ARAÚJO, D. G. A gestão dos dados de pesquisa no âmbito da comunidade dos pesquisadores vinculados aos programas de pós-

graduação brasileiros na área da Ciência da Informação: desvendando as práticas e percepções associadas ao uso e reuso de dados. **Liinc em Revista**, Rio de Janeiro, v. 15, n. 2, p. 5-31, nov. 2019. DOI: <https://doi.org/10.18617/liinc.v15i2.4683>. Disponível em: <http://revista.ibict.br/liinc/article/view/4683/4327>. Acesso em: 14 out. 2020.

DIAS, G. A.; ANJOS, R. L.; RODRIGUES, A. A. Os princípios FAIR: viabilizando o reuso de dados científicos. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados científicos**: perspectivas e desafios. João Pessoa: Ed. UFPB, 2019. p. 177-187. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/view/359/508/2949-1>. Acesso em: 29 jun. 2020.

DIGITAL CURATION CENTRE. **About the DCC**. c2019a. Disponível em: <http://www.dcc.ac.uk/about-us>. Acesso em: 12 set. 2019.

DIGITAL CURATION CENTRE. **Data management plans**. c2019b. Disponível em: <http://www.dcc.ac.uk/resources/data-management-plans>. Acesso em: 16 set. 2019

DIGITAL CURATION CENTRE. **DCC curation lifecycle model**. c2019c. Disponível em: <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. Acesso em: 12 set. 2019.

DIGITAL CURATION CENTRE. **History of the DCC**. c2019d. Disponível em: <http://www.dcc.ac.uk/about-us/history-dcc/history-dcc>. Acesso em: 12 set. 2019.

DIGITAL CURATION CENTRE. **What is digital curation?**. c2019e. Disponível em: <http://www.dcc.ac.uk/digital-curation/what-digital-curation>. Acesso em: 12 set. 2019.

DUBLIN CORE METADATA INITIATIVE. **[Arquivo Readme.md]**. 2020. Disponível em: <https://github.com/dcmi/dcap>. Acesso em: 21 out. 2020.

DUBLIN CORE METADATA INITIATIVE. **About DCMI**. c2020a. Disponível em: <https://www.dublincore.org/about/>. Acesso em: 17 mar. 2020.

DUBLIN CORE METADATA INITIATIVE. **Metadata basics**. c2020b. Disponível em: <https://www.dublincore.org/resources/metadata-basics/>. Acesso em: 17 mar. 2020.

DUERST, M.; SUIGNARD, M. **Internationalized Resource Identifiers (IRIs)**. [S.l.]: RFC Editor, 2005. Disponível em: <https://www.rfc-editor.org/info/rfc3987>. Acesso em: 3 jul. 2018.

ETSIN. **[Página inicial]**. Disponível em: <https://etsin.fairdata.fi/>. Acesso em: 23 dez. 2020.

EUROPEAN DATA PORTAL. **About the European Data Portal**. 2020a. Disponível em: <https://www.europeandataportal.eu/en/about/european-data-portal>. Acesso em: 23 dez. 2020.

EUROPEAN DATA PORTAL. [Registro do dataset 'Etat du trafic en temps réel']. 2020b. Disponível em: <https://www.europeandataportal.eu/data/datasets/5caaf5ee9ce2e75d0c8c381a?locale=en>. Acesso em: 23 dez. 2020.

EUROPEAN UNION OPEN DATA PORTAL. **About**. [2020?]a. Disponível em: <https://data.europa.eu/euodp/en/about>. Acesso em: 28 dez. 2020.

EUROPEAN UNION OPEN DATA PORTAL. **Linked Data**. [2020?]b. Disponível em: <https://data.europa.eu/euodp/en/linked-data>. Acesso em: 28 dez. 2020.

FAIRDATA IDA. **Welcome to Fairdata IDA**. [2020?]. Disponível em: <https://ida.fairdata.fi/login>. Acesso em: 28 dez. 2020.

FALLSIDE, D.; WALMSLEY, P. (ed.). **XML Schema part 0**: primer second edition. 2004. Disponível em: <http://www.w3.org/TR/2004/REC-xmlschema-0-20041028/>. Acesso em: 22 abr. 2020.

FARNEL, S.; SHIRI, A. Metadata for research data: current practices and trends. *In*: DCMI INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2014, Austin. **Proceedings** [...]. Austin, TX: DCMI, 2014. p. 74-82. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3714/1937>. Acesso em: 22 jun. 2020.

FERREIRA, J. A.; SANTOS, P. L. V. A. C. O modelo de dados Resourc 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/15436/9681>. Acesso em: 15 out. 2020.

FERREIRA, N. S. A. As pesquisas denominadas “estado da arte”. **Educação & Sociedade**, v. 23, n. 79, p. 257-272, ago. 2002. DOI: <http://dx.doi.org/10.1590/S0101-73302002000300013>. Disponível em: <https://www.scielo.br/pdf/es/v23n79/10857.pdf>. Acesso em: 16 out. 2018.

FONSECA, D. L. S.; MONTEIRO, E. W. C. Encontrabilidade da informação e os princípios FAIR: meios alternativos de gestão de dados abertos. **Ciência da Informação**, Brasília, v. 48, n. 3 (Supl.), p. 507-508, set./dez. 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4971>. Acesso em: 15 out. 2020.

FORCE11. **The FAIR data principles**. c2017. Disponível em: <https://www.force11.org/group/fairgroup/fairprinciples>. Acesso em: 17 set. 2019.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Auxílio à pesquisa projeto temático**. [2019?]. Disponível em: <http://www.fapesp.br/176>. Acesso em: 14 set. 2019.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Código de boas práticas científicas**. 2014. Disponível em:

[www.fapesp.br/boaspraticas/FAPESP-Codigo de Boas Praticas Cientificas 2014.pdf](http://www.fapesp.br/boaspraticas/FAPESP-Codigo%20de%20Boas%20Praticas%20Cientificas%202014.pdf). Acesso em: 13 nov. 2018.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Plano de gestão de dados – FAPESP**. [2018?]. Disponível em: <http://fapesp.br/gestaodedados/>. Acesso em: 25 set. 2018.

FUNDAÇÃO DE AMPARO À PESQUISA DO ESTADO DE SÃO PAULO. **Planos de gestão de dados se incorporam a projetos de pesquisa no Brasil**. 2017. Disponível em: <https://revistapesquisa.fapesp.br/2017/10/25/planos-de-gestao-de-dados-se-incorporam-a-projetos-de-pesquisa-no-brasil/>. Acesso em: 14 set. 2019.

GEONAMES. **About GeoNames**. [2020?]. Disponível em: <https://www.geonames.org/about.html>. Acesso em: 15 abr. 2020.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: ATLAS S.A., 2002.

GILLILAND, A. J. Setting the Stage. *In*: BACA, Murtha (Ed.). **Introduction to Metadata**. 3rd ed. Los Angeles: Getty P., 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/setting-the-stage/>. Acesso em: 15 jan. 2020.

GO FAIR. **FAIR principles**. [2019?]. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 24 set. 2019.

GOVDATA. **Das portal**. 2016. Disponível em: <https://www.govdata.de/web/guest/hilfe>. Acesso em: 28 dez. 2020.

GOVDATA. **[Registro do dataset ‘Bevölkerung nach Altersjahren und Geschlecht’]**. 2020. Disponível em: <https://www.govdata.de/web/guest/suchen/-/details/bevolkerung-nach-altersjahren-und-geschlecht>. Acesso em: 28 dez. 2020.

GRAY, J. *et al.* Online scientific data curation, publication, and archiving. (A. S. Szalay, Ed.). *In*: ASTRONOMICAL TELESCOPES AND INSTRUMENTATION, 2002, Waikoloa. **Proceedings** [...]. Redmond: SPIE, 2002. p. 103-107. Disponível em: <https://www.microsoft.com/en-us/research/publication/online-scientific-data-curation-publication-and-archiving/>. Acesso em: 7 jul. 2018.

HARRIS, S.; SEABORNE, A. (ed.). **SPARQL 1.1 query language**. 2013. Disponível em: <http://www.w3.org/TR/2013/REC-sparql11-query-20130321/>. Acesso em: 07 jan. 2020.

HEATH, T. **Linked Data**. [201-?]. Disponível em: <http://linkeddata.org/home>. Acesso em: 26 ago. 2019.

HEATH, T.; BIZER, C. **Linked Data: evolving the Web into a global space**. [S.l.]: Morgan & Claypool, 2011. (Synthesis Lectures on the Semantic Web: Theory and Technology, v. 1, n.1). Versão HTML. Disponível em: <http://linkeddatabook.com/editions/1.0/>. Acesso em: 23 mar. 2020.

HEERY, R.; PATEL, M. Application profiles: mixing and matching metadata schemas. **Ariadne**, n. 25, 2000. Disponível em: <http://www.ariadne.ac.uk/issue/25/app-profiles/>. Acesso em 30 mar. 2020.

HENNING, P. C. GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 389-412, maio/ago. 2019. DOI: <http://dx.doi.org/10.19132/1808-5245252.p.389-412>. Disponível em: <https://seer.ufrgs.br/EmQuestao/article/view/84753/52667>. Acesso em: 15 out. 2020.

HEY, T.; TANSLEY, S.; TOLLE, K. (Ed.). **The fourth paradigm**: data-intensive scientific discovery. Microsoft Research: Redmond, 2009. Disponível em: https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf. Acesso em: 02 jul. 2020.

HORROCKS, I.; PATEL-SCHNEIDER, P. F. **Three theses of representation in the Semantic Web**. 2003. Disponível em: <http://www2003.org/cdrom/papers/refereed/p050/p50-horrocks.html>. Acesso em: 21 out. 2020.

IANNELLA, R. *et al.* (ed.). **ODRL vocabulary and expression 2.2**. 2018. Disponível em: <https://www.w3.org/TR/2018/REC-odrl-vocab-20180215/>. Acesso em: 11 maio 2020.

IANNELLA, R.; MCKINNEY, J. **vCard Ontology - for describing people and organizations**. 2014. Disponível em: <http://www.w3.org/TR/2014/NOTE-vcards-rdf-20140522/><https://www.w3.org/TR/vcard-rdf/>. Acesso em: 10 ago. 2018.

ICOS CARBON PORTAL. **[Registro do dataset 'ICOS ATC NRT CO2 growing time series from Zeppelin']**. 2020. Disponível em: <https://data.icos-cp.eu/portal/#%7B%22route%22%3A%22metadata%22%2C%22id%22%3A%22EeHJrPwgCmHTKUdswQSqX559%22%7D>. Acesso em: 28 dez. 2020.

IDEHEN, K. U. **Semantic Web layer cake tweak, explained**. 2017. Disponível em: <https://medium.com/openlink-software-blog/semantic-web-layer-cake-tweak-explained-6ba5c6ac3fab>. Acesso em: 21 out. 2020.

INTEGRATED CARBON OBSERVATION SYSTEM. **About and contacts**. [2020?]a. Disponível em: <https://www.icos-cp.eu/about>. Acesso em: 28 dez. 2020.

INTEGRATED CARBON OBSERVATION SYSTEM. **ICOS Carbon Portal**. [2020?]b. Disponível em: <https://www.icos-cp.eu/observations/carbon-portal>. Acesso em: 28 dez. 2020.

ISAAC, A.; PHIPPS, J.; RUBIN, D. (ed.). **SKOS use cases and requirements**. 2009. Disponível em: <https://www.w3.org/TR/2009/NOTE-skos-ucr-20090818/>. Acesso em: 29 abr. 2020.

ISAAC, A.; SUMMERS, E. (ed.). **SKOS Simple Knowledge Organization System primer**. 2009. Disponível em: <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>. Acesso em: 9 ago. 2018.

ISOTANI, S.; BITTENCOURT, I. I. **Dados abertos conectados**. São Paulo: Novatec Ed., 2015.

JOINT RESEARCH CENTRE DATA CATALOGUE. **About the JRC Data Catalogue**. 2020. Disponível em: <https://data.jrc.ec.europa.eu/about>. Acesso em: 29 dez. 2020.

JOINT RESEARCH CENTRE DATA CATALOGUE. **[Registro do dataset '01_JRC-EU-TIMES Full model']**. 2019. Disponível em: <https://data.jrc.ec.europa.eu/dataset/8141a398-41a8-42fa-81a4-5b825a51761b>. Acesso em: 29 dez. 2020.

KIM, Y.; STANTON, J. M. Institutional and individual influences on scientists' data sharing behaviors: a multilevel analysis. *In*: AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY ANNUAL MEETING, 76., 2013, Montreal. **Proceedings** [...]. [S.l.]: ASIS&T, 2013. Disponível em: <https://doi.org/10.1002/meet.14505001093>. Acesso em: 18 jun. 2020.

KUHN, T. S. Introdução: um papel para a história. *In*: KUHN, T. S. **A estrutura das revoluções científicas**. 9. ed. São Paulo: Perspectiva, 2006. p. 19-28.

LAENDER, A. H. F. *et al.* **Abertura e Gestão de dados**: desafios para a Ciência brasileira. [S.l.]: Academia Brasileira de Ciências, 2020. Disponível em: <http://www.abc.org.br/wp-content/uploads/2020/09/ABC-Abertura-e-Gest%C3%A3o-de-Dados-desafios-para-a-ci%C3%Aancia-brasileira.pdf>. Acesso em: 07 out. 2020.

LANDIM FILHO, R. Conceito e objeto em Tomás de Aquino. **Analytica**, Rio de Janeiro, v. 14, n. 2, p. 65-88, 2010. Disponível em: <https://revistas.ufrj.br/index.php/analytica/article/view/598/554>. Acesso em: 16 jan. 2020.

LATOUR, B. Introdução: abrindo a caixa-preta de Pandora. *In*: LATOUR, B. **Ciência em ação**: como seguir cientistas e engenheiros sociedade afora. São Paulo: Editora UNESP, 2000. p. 11-36.

LEBO, T.; SAHOO, S.; MCCGUINNESS, D. (ed.). **PROV-O**: the PROV Ontology. 2013. Disponível em: <http://www.w3.org/TR/2013/REC-prov-o-20130430/>. Acesso em: 15 jul. 2020.

LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. (ed.). **Data on the Web best practices**. 2017. Disponível em: <https://www.w3.org/TR/2017/REC-dwbp-20170131/>. Acesso em: 26 mar. 2020.

MALTA, M. C.; BAPTISTA, A. A. A method for the development of Dublin Core Application Profiles (Me4DCAP V0.2): detailed description. *In*: DCMI INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2013, Lisbon. **Proceedings** [...]. Lisbon: DCMI, 2013. p. 90-103.

Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3674/1897>. Acesso em: 16 out. 2020.

MALTA, M. C.; BAPTISTA, A. A. A panoramic view on metadata application profiles of the last decade. **International Journal of Metadata, Semantics and Ontologies**, Geneva, v. 9, n. 1, p. 58-73, 2014. DOI: <https://doi.org/10.1504/IJMSO.2014.059124>. Disponível em: https://repositorium.sdum.uminho.pt/bitstream/1822/50751/1/Malta%26Baptista_aPanoramicView.pdf. Acesso em: 16 out. 2020.

MARCONDES, C. H. Representação e economia da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 61-70, jan./abr. 2001. Disponível em: <http://revista.ibict.br/ciinf/article/view/939/976>. Acesso em: 16 jan. 2020.

MCCARRON, S. (Ed.). **XHTML+RDFa 1.1 - Third Edition**. 2015. Disponível em: <http://www.w3.org/TR/2015/REC-xhtml-rdfa-20150317/>. Acesso em: 04 jan. 2021.

MCCRAE, J.P. *et al.* **The Linked Open Data cloud**. [2021?]. Disponível em: <https://lod-cloud.net/>. Acesso em: 04 fev. 2021.

MCKEN, K. *et al.* **Research360: data in the research lifecycle**. 2012. Disponível em: <https://researchportal.bath.ac.uk/en/publications/research360-data-in-the-research-lifecycle>. Acesso em: 29 jun. 2020.

MILES, A.; BECHHOFER, S. (ed.). **SKOS Simple Knowledge Organization System reference**. 2009. Disponível em: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>. Acesso em: 9 ago. 2018.

MINISTERIO DE EDUCACIÓN. **Acerca de**: repositorio de datos del Ministerio de Educación. [2021?]. Disponível em: <http://datos.minedu.gob.pe/acerca-de>. Acesso em: 04 jan. 2021.

MINISTERIO DE EDUCACIÓN. **[Registro do dataset 'Indicadores de Educación Básica Alternativa']**. 2019. Disponível em: <http://datos.minedu.gob.pe/dataset/indicadores-de-educaci%C3%B3n-b%C3%A1sica-alternativa>. Acesso em: 04 jan. 2021.

MINUSI, S. G. *et al.* **Considerações sobre estado da arte, levantamento bibliográfico e pesquisa bibliográfica**: relações e limites. *Revista Gestão Universitária*, 2018. Disponível em: <http://gestaouniversitaria.com.br/artigos/consideracoes-sobre-estado-da-arte-levantamento-bibliografico-e-pesquisa-bibliografica-relacoes-e-limites>. Acesso em: 26 ago. 2019.

MONTEIRO, E. C. S. A.; SANT'ANA, R. C. G. Repositórios de dados científicos na infraestrutura de pesquisa: adoção dos princípios FAIR. **Ciência da Informação**, Brasília, v. 48, n. 3 (Supl.), p. 347-353, set./dez. 2019. Disponível em: <http://revista.ibict.br/ciinf/article/view/4878/4448>. Acesso em: 15 out. 2020.

- MONTEIRO, G.; LUCAS, E. R. O. Dados científicos abertos: políticas de financiadores da pesquisa brasileira. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 20., 2019, Florianópolis. **Anais [...]**. Florianópolis:UFSC, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/881/931>. Acesso em: 04 jun. 2020.
- MOREAU, L.; MISSIER, P. (ed.). **PROV-DM**: the PROV Data Model. 2013a. Disponível em: <http://www.w3.org/TR/2013/REC-prov-dm-20130430/>. Acesso em: 15 jul. 2020.
- MOREAU, L.; MISSIER, P. (ed.). **PROV-N**: the provenance notation. 2013b. Disponível em: <http://www.w3.org/TR/2013/REC-prov-n-20130430/>. Acesso em: 15 jul. 2020.
- MORENO, F. P. Repositório de dados de pesquisa na Espanha: breve análise. **Encontros Bibli**, v. 23, n. 53, p. 52-63, set./dez. 2018. DOI: 10.5007/1518-2924.2018v23n53p52. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2018v23n53p52>. Acesso em: 14 out. 2020.
- NHACUONGUE, J. A.; FERNEDA, E. O Campo da Ciência da Informação: contribuições, desafios e perspectivas. **Perspectivas em Ciência da Informação**, [s.l.], v. 20, n. 2, p. 3-18, jun. 2015. DOI: <http://dx.doi.org/10.1590/1981-5344/1932>. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1932/1591>. Acesso em: 31 mar. 2019.
- NÓBREGA-TERRIEN, S. M.; TERRIEN, J. Trabalhos científicos e o estado da questão. **Estudos em Avaliação Educacional**, v. 15, n. 30, p. 5-16, 30 dez. 2004. DOI: <http://dx.doi.org/10.18222/eae153020042148>. Disponível em: <http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/2148/2105>. Acesso em: 11 set. 2018.
- NORWEGIAN POLAR DATA CENTRE. **[Registro do dataset 'Elevation change and mass budget of Ross and Filchner-Ronne ice shelves, Antarctica']**. 2016. Disponível em: <https://data.npolar.no/dataset/cae21585-1e7e-4f86-b6be-a8b13696d402>. Acesso em: 29 dez. 2020.
- NORWEGIAN POLAR INSTITUTE. **What we do**. [2020?]. Disponível em: <https://www.npolar.no/en/about-us/#toggle-id-2>. Acesso em: 29 dez. 2020.
- NOWACK, B. **The Semantic Web**: not a piece of cake. 2009. Disponível em: <http://bnode.de/blog/2009/07/08/the-semantic-web-not-a-piece-of-cake>. Acesso em: 29 ago. 2019.
- OLIVEIRA, A. C. S. et al. Adopting the Dublin Core standard for describing open scientific data: the e-Quilt prototype experiment. *In: DCMI INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS*, 2015, São

Paulo. **Proceedings** [...]. São Paulo: DCMI, 2015. p. 223-225. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3788>. Acesso em: 02 jul. 2020.

OPEN DATA KINGSTON. **Data**. [2021?]a. Disponível em: <https://opendatakingston.cityofkingston.ca/explore/?sort=modified>. Acesso em: 04 jan. 2021.

OPEN DATA KINGSTON. **Welcome to the Open Data Kingston Portal**. [2021?]b. Disponível em: <https://opendatakingston.cityofkingston.ca/pages/welcome/>. Acesso em: 04 jan. 2021.

OPEN GOVERNMENT CANADA. **Open Data**. 2020. Disponível em: <https://open.canada.ca/en/open-data>. Acesso em: 04 jan. 2021.

OPEN GOVERNMENT CANADA. **[Registro do dataset 'Proactive Disclosure - Acts of Founded Wrongdoing']**. 2021. Disponível em: <https://open.canada.ca/data/en/dataset/6e75f19c-d19d-48aa-984e-609c8d9bc403>. Acesso em: 04 jan. 2021.

OPEN KNOWLEDGE FOUNDATION. **What is open data?**. In: Open data handbook. [S.l.: s.n.], [2019?]. Disponível em: <http://opendatahandbook.org/guide/en/what-is-open-data/>. Acesso em: 28 ago. 2019.

OPENDATA. **[Página inicial]**. [2021?]. Disponível em: <https://opendata.socrata.com/>. Acesso em: 04 fev. 2021.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **OECD principles and guidelines for access to research data from public funding**. 2007. Disponível em: <http://www.oecd.org/sti/inno/38500813.pdf>. Acesso em: 02 abr. 2019.

PATTUELLI, M. C. Personal name vocabularies as linked open data: a case study of jazz artist names. **Journal of Information Science**, v. 38, n. 6, p. 558-565, out. 2012. DOI: <https://doi.org/10.1177%2F0165551512455989>.

PERREAULT, S. **vCard format specification**. 2011. Disponível em: <https://tools.ietf.org/html/rfc6350#section-1>. Acesso em: 10 ago. 2018.

PESET, F.; FERRER-SAPENA, A.; SUBIRATS-COLL, I. Open data y linked open data: su impacto en el área de bibliotecas y documentación. **El profesional de la información**, v. 20, n. 2, p. 165-173, marzo-abr. 2011. Disponível em: <http://hdl.handle.net/10251/29837>. Acesso em: 29 ago. 2019.

PITTA, G. B. B.; CASTRO, A. A. A pesquisa científica. **Jornal Vascular Brasileiro**, Porto Alegre, v. 5, n. 4, p. 243-244, dez. 2006. DOI: <http://dx.doi.org/10.1590/S1677-54492006000400001>. Disponível em: <http://www.scielo.br/pdf/jvb/v5n4/v5n4a01.pdf>. Acesso em: 11 out. 2018.

POMERANTZ, J. **Metadata**. Cambridge: The MIT Press, 2015.

POWELL, A.; JOHNSTON, P. **Guidelines for implementing Dublin Core in XML**. 2003. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dc-xml-guidelines/>. Acesso em: 05 maio 2020.

POWELL, A. *et al.* **DCMI Abstract Model**. 2007. Disponível em: <http://dublincore.org/documents/abstract-model/>. Acesso em: 13 mar. 2020.

QIN, J.; LI, K. How portable are the metadata standards for scientific data? A proposal for a metadata infrastructure. *In: DCMI INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2013, Lisboa. Proceedings* [...]. Lisboa: DCMI, 2013. Disponível em: <https://dcpapers.dublincore.org/pubs/article/view/3670/1893>. Acesso em: 19 jun. 2020.

RAMALHO, R. A. S. Análise do modelo de dados SKOS: sistema de organização do conhecimento simples para a Web. **ITEC**, Marília/João Pessoa, v. 2, n. 1, p. 66-79, jan./jul. 2015. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/itec/article/view/25995/14680>. Acesso em: 15 out. 2020.

RAMALHO, R. A. S. O modelo de dados SKOS: novas perspectivas no âmbito da representação do conhecimento. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Bahia. Anais* [...]. Salvador: PPGCI/UFBA, 2016. Disponível em: <http://www.ufpb.br/evento/index.php/enancib2016/enancib2016/paper/view/4078>. Acesso em: 15 out. 2020.

RAUTENBERG, S. *et al.* **Guia prático para publicação de dados abertos conectados na web**. Curitiba: Appris, 2018. (Educação, tecnologias e transdisciplinaridade).

RDPCIDAT. **About**. [2021?]. Disponível em: <https://rdpcidat.rub.de/about>. Acesso em: 04 jan. 2021.

RDPCIDAT. **[Registro do dataset 'Plasma-driven in situ production of hydrogen peroxide for biocatalysis']**. 2020. Disponível em: <https://rdpcidat.rub.de/dataset/plasma%E2%80%90driven-%E2%80%85situ-production-hydrogen-peroxide-biocatalysis>. Acesso em: 04 jan. 2021.

REGISTRY OF RESEARCH DATA REPOSITORIES. **About**. [2020?]. Disponível em: <https://www.re3data.org/about>. Acesso em: 21 dez. 2020.

REGISTRY OF RESEARCH DATA REPOSITORIES. **Biological and Chemical Oceanography Data Management Office**. 2020a. Disponível em: <https://www.re3data.org/repository/r3d10000012>. Acesso em: 02 fev. 2021.

REGISTRY OF RESEARCH DATA REPOSITORIES. **[Busca por 'data catalog vocabulary']**. 2020b. Disponível em: <https://www.re3data.org/search?query=data%20catalog%20vocabulary&metadataSt>

[andards%5B%5D=DCAT%20-%20Data%20Catalog%20Vocabulary](#). Acesso em: 28 dez. 2020.

REGISTRY OF RESEARCH DATA REPOSITORIES. **ICOS Carbon Portal**. 2019. Disponível em: <https://www.re3data.org/repository/r3d100012203>. Acesso em: 28 dez. 2020.

REGISTRY OF RESEARCH DATA REPOSITORIES. **Norwegian Polar Data Centre**. 2020c. Disponível em: <https://www.re3data.org/repository/r3d100012291>. Acesso em: 29 dez. 2020.

REGISTRY OF RESEARCH DATA REPOSITORIES. **Open Data by Socrata**. 2018. Disponível em: <https://www.re3data.org/repository/r3d100011777>. Acesso em 29 dez. 2020.

RÉPERTOIRE INTERNATIONAL DES SOURCES MUSICALES. **About**. [2021?]a. Disponível em: <https://opac.rism.info/main-menu-/kachelmenu/about>. Acesso em: 04 jan. 2021.

RÉPERTOIRE INTERNATIONAL DES SOURCES MUSICALES. [**Registro do recurso 'Tröstet, tröstet ...'**]. [2021?]b. Disponível em: https://opac.rism.info/metaopac/singleHit.do?methodToCall=showHit&curPos=1&identifier=251_SOLR_SERVER_214907328. Acesso em: 04 jan. 2021.

RÉPERTOIRE INTERNATIONAL DES SOURCES MUSICALES. **SPARQL**. [2021?]c. Disponível em: <https://opac.rism.info/sparql-endpoint>. Acesso em: 04 jan. 2021.

RÉPERTOIRE INTERNATIONAL DES SOURCES MUSICALES. **Data**: Open Data, SPARQL, etc. 2020. Disponível em: <https://opac.rism.info/main-menu-/kachelmenu/data>. Acesso em: 04 jan. 2021.

RESEARCH DATA ALLIANCE. **23 things**: libraries for research data. 2016. Disponível em: https://rd-alliance.org/system/files/documents/23Things_Libraries_For_Data_Management.pdf. Acesso em: 14 out. 2020.

RESEARCH INFORMATION NETWORK. **To share or not to share**: publication and quality assurance of research data outputs. [S.l.: s.n.], 2008. Disponível em: https://eprints.soton.ac.uk/266742/1/Published_report_-_main_-_final.pdf. Acesso em: 09 set. 2018.

RESENDE, L. C.; BAX, M. P. Curadoria digital de dados científicos: o cenário brasileiro na Ciência da Informação. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. **Anais [...]**. Florianópolis: UFSC, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/983/625>. Acesso em: 14 out. 2020.

RIBEIRO, C. J. S. Modelo de maturidade para repositórios digitais: um caminho para sua adoção na gestão de dados de pesquisa. **Liinc em Revista**, Rio de Janeiro, v.

15, n. 2, p. 224-243, nov. 2019. DOI: <https://doi.org/10.18617/liinc.v15i2.4816>. Disponível em: <http://revista.ibict.br/liinc/article/view/4816/4317>. Acesso em: 14 out. 2020.

RILEY, J. **Understanding metadata**: what is metadata, and what is it for?. [S.l.]: National Information Standards Organization (NISO), 2017. Disponível em: https://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Metadata.pdf. Acesso em: 07 jan. 2020.

ROCHA, L. L.; SALES, L. F.; SAYÃO, L. F. **Descrever para preservar**: metadados como ferramenta para gestão de dados de pesquisa. In: PINHO, A. P.; GUIMARÃES, J. A. C. (Org.). *Memória, tecnologia e cultura na organização do conhecimento*. Recife: Ed. UFPE, 2017. p. 194-201. (Estudos avançados em organização do conhecimento, 4). Disponível em: <http://isko-brasil.org.br/wp-content/uploads/2013/02/livro-ISKO-2017.pdf>. Acesso em: 04 jun. 2020.

ROZSA, V.; DUTRA, M. L.; NHACUONGUE, J. A. Linked Open Data no contexto acadêmico: identificação e análise de vocabulários utilizados na academia e na pesquisa científica. **Brazilian Journal of Information Studies: Research Trends**, v. 11, n. 3, 2017. p.34-52. DOI: <https://doi.org/10.36311/1981-1640.2017.v11n3.05.p34>. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/6780/4651>. Acesso em: 07 ago. 2020.

SALES, L. F.; SAYÃO, L. F. Uma proposta de taxonomia para dados de pesquisa. **Conhecimento em ação**, Rio de Janeiro, v. 4, n. 1, jan./jun. 2019. DOI: <https://doi.org/10.47681/rca.v4i1.26337>. Disponível em: <https://revistas.ufrj.br/index.php/rca/article/view/26337/14573>. Acesso em: 22 out. 2020.

SANT'ANA, R. C. G. Campo informacional resultante da interação de ciclos de vida dos dados. In: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: Ed. UFPB, 2019. p. 13-31. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/view/359/508/2949-1>. Acesso em: 29 jun. 2020.

SANTARÉM SEGUNDO, J. E. Web Semântica: introdução a recuperação de dados usando SPARQL. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte. **Anais [...]**. Belo Horizonte: ECI/UFMG, 2014. p. 3863-3882. Disponível em: <http://enancib2014.eci.ufmg.br/documentos/anais/anais-gt8>. Acesso em: 28 ago. 2019.

SANTOS, P. L. V.; SANT'ANA, R. C. G. Camadas de representação de dados e suas especificidades no cenário científico. In: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: Ed. UFPB, 2019. p. 53-66. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/view/359/508/2949-1>. Acesso em: 29 jun. 2020.

SANTOS, P. L.V. A.C.; SANT'ANA, R. C. G. Dado e granularidade na perspectiva da informação e tecnologia: uma interpretação pela Ciência da Informação. **Ciência da Informação**, Brasília, v. 42, n. 2, jan. 2015. DOI:

<https://doi.org/10.18225/ci.inf..v42i2.1382>. Disponível em:

<http://revista.ibict.br/ciinf/article/view/1382>. Acesso em: 30 jun. 2020.

SANTOS, P. L. V. A. C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia BEAM.

Informação & Informação, Londrina, v. 19, n. 1, p. 146-163, jan./abr. 2014. DOI:

<http://dx.doi.org/10.5433/1981-8920.2014v19n1p146>. Disponível em:

<http://www.uel.br/revistas/uel/index.php/informacao/article/view/15251/14211>.

Acesso em: 21 dez. 2020.

SANTOS, P. L. V. A. C.; VIDOTTI, S. A. B. G. Perspectivismo e tecnologias de informação e comunicação: acréscimos à Ciência da Informação. **DataGramZero**, v.

10, n. 3, 2009. Disponível em: <https://brapci.inf.br/index.php/article/download/11618>.

Acesso em: 02 jul. 2020.

SARACEVIC.T. **Information science**. In: Marcia J. Bates and Mary Niles Maack

(Eds.) *Encyclopedia of Library and Information Science*. New York: Taylor & Francis.

p. 2570-2586, 2009. Disponível em:

<https://tefkos.cominfo.rutgers.edu/SaracevicInformationScienceELIS2009.pdf>.

Acesso em: 02 maio 2019.

SAYÃO, L. F.; SALES, L. F. A ciência invisível: os dados da cauda longa da pesquisa. In: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados científicos:**

perspectivas e desafios. João Pessoa: Ed. UFPB, 2019. p. 33-52. Disponível em:

<http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/view/359/508/2949-1>. Acesso em: 29 jun. 2020.

SAYÃO, L. F.; SALES, L. F. Curadoria digital e dados de pesquisa. **AtoZ: novas**

práticas em informação e conhecimento, v. 5, n. 2, p. 67-71, jul./dez. 2016. DOI:

<http://dx.doi.org/10.5380/atoz.v5i2.49708>. Disponível em:

<https://revistas.ufpr.br/atoz/article/view/49708/30161>. Acesso em: 14 set. 2019.

SAYÃO, L. F.; SALES, L. F. **Guia de gestão de dados de pesquisa para**

bibliotecários e pesquisadores. Rio de Janeiro: CNEN, 2015. Disponível em:

http://www.cnen.gov.br/images/CIN/PDFs/GUIA_DE_DADOS_DE_PESQUISA.pdf.

Acesso em: 13 set. 2019.

SCHREIBER, G.; RAIMOND, Y. (ed.). **RDF 1.1 primer**. 2014. Disponível em:

<https://www.w3.org/TR/rdf11-primer/>. Acesso em: 3 jul. 2018.

SEMELER, A. R.; PINTO, A. L. Os diferentes conceitos de dados de pesquisa na abordagem da biblioteconomia de dados. **Ciência da Informação**, Brasília, DF, v. 48, n. 1, p. 113-129, jan./abr. 2019. Disponível em:

<http://revista.ibict.br/ciinf/article/view/4461/4102>. Acesso em: 10 mar. 2020.

SHIRADO, L. Y.; RAMALHO, R. A. S. A inovação e as tecnologias semânticas no processo de representação e organização da informação. In: COLÓQUIO EM

ORGANIZAÇÃO, ACESSO E APROPRIAÇÃO DA INFORMAÇÃO E DO CONHECIMENTO, 4., 2019, Londrina. **Anais** [...]. Londrina: UEL, 2019. Disponível em:

<http://www.uel.br/eventos/cinf/index.php/coaic2019/coaic2019/paper/viewFile/616/437>. Acesso em: 15 out. 2020.

SILVA, J. L. C.; GOMES, H. F. Conceitos de informação na Ciência da Informação: percepções analíticas, proposições e categorizações. **Informação & Sociedade: Estudos**, João Pessoa, v. 25, n. 1, p. 145-157, jan./abr. 2015. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/ies/article/view/145/13200>. Acesso em: 02 jul. 2020.

SILVA, L. C.; SANTAREM SEGUNDO, J. E.; SILVA, M. F. Princípios de FAIR e melhores práticas do linked data na publicação de dados de pesquisa. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 19., 2018, Londrina. **Anais** [...]. Londrina: UEL, 2018. p. 5195-5215.

SILVA SEGUNDO, S. J.; ARAÚJO, W. J. Curadoria e ciclo de vida dos dados. *In*: DIAS, G. A.; OLIVEIRA, B. M. J. F. (org.). **Dados científicos: perspectivas e desafios**. João Pessoa: Ed. UFPB, 2019. p. 113-152. Disponível em: <http://www.editora.ufpb.br/sistema/press5/index.php/UFPB/catalog/view/359/508/2949-1>. Acesso em: 29 jun. 2020.

SIMIONATO, A. C. Mapeamento dos metadados para dados científicos. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. **Anais** [...]. Marília: UNESP, 2017. Disponível em: <http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/563/874>. Acesso em: 04 jun. 2020.

SIMIONATO, A. C. **Representação, acesso, uso e reuso da imagem digital**. 141f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista – UNESP, Marília, 2012. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/Simionato%20A.C._mestrado_C.I._2012.pdf. Acesso em: 06 out. 2020.

SMIT, E.; VAN DER HOEVEN, J.; GIARETTA, D. Avoiding a Digital Dark Age for data: why publishers should care about digital preservation. **Learned Publishing**, v. 24, n. 1, P. 35-49, jan. 2011. DOI: <https://doi.org/10.1087/20110107>. Disponível em: <https://onlinelibrary.wiley.com/doi/epdf/10.1087/20110107>. Acesso em: 12 set. 2019.

SMIT, J. W. A informação na Ciência da Informação. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 3, n. 2, p. 84-101, jul./dez. 2012. Disponível em: <https://doi.org/10.11606/issn.2178-2075.v3i2p84-101>. Acesso em: 19 abr. 2019.

SMITH, V. S. Data publication: towards a database of everything. **BMC Research Notes**, v. 2, n. 1, p. 113, jun. 2009. DOI: <https://doi.org/10.1186/1756-0500-2-113>. Disponível em: <https://bmcresearchnotes.biomedcentral.com/articles/10.1186/1756-0500-2-113>. Acesso em: 10 ago. 2018.

STRASSER, C. *et al.* **Primer on data management**: what you always wanted to know. Oakland: California Digital Library, 2012. DOI: <https://doi.org/doi:10.5060/D2251G48>. Disponível em: <https://escholarship.org/uc/item/7tf5q7n3>. Acesso em: 16 set. 2019.

TÁLAMO, M. F.; SMIT, J.W. Ciência da informação: pensamento informacional e integração disciplinar. **Brazilian Journal of Information Science**, v. 1, n. 1, p. 33-57, jan./jun. 2007. DOI: <https://doi.org/10.36311/1981-1640.2007.v1n1.03.p33>. Disponível em: <http://www2.marilia.unesp.br/revistas/index.php/bjis/article/view/30/29>. Acesso em: 02 jul. 2020.

TARTAROTTI, R. C. D.; DAL'EVEDOVE, P. R.; FUJITA, M. S. L. Atuação bibliotecária em repositórios de dados de pesquisa: da perspectiva da entrega à perspectiva do acesso. *In*: BARROS, Thiago Henrique Bragato; TOGNOLI, Natalia Bolfarini (org.). **Organização do conhecimento responsável**: promovendo sociedades democráticas e inclusivas. Belém: Ed. Da UFPA: ISKO-Brasil, 2019. P. 451-458. Disponível em: <http://isko-brasil.org.br/wp-content/uploads/2019/09/LIVRO-ISKO-BRASIL-EDI%C3%87%C3%83O-BEL%C3%89M.pdf>. Acesso em: 10 mar. 2020.

TEMESIO, S. Metadatos de datos abiertos: estudio de caso en Uruguay. **BiD: textos universitarios de biblioteconomía e documentación**, n. 34, jun., 2015. DOI: <http://dx.doi.org/10.1344/BiD2015.34.18>. Disponível em: <http://bid.ub.edu/es/34/temesio.htm>. Acesso em: 14 jul. 2020.

TENOPIR, C. *et al.* Research data services in 178merican and north 178merican libraries: current offerings and plans for the future. *In*: ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY, 2016, Copenhagen. **Proceedings** [...]. [S.l.]: ASIS&T, 2016. Disponível em: <https://doi.org/10.1002/pra2.2016.14505301129>. Acesso em: 18 jun. 2020.

THE CASRAI DICTIONARY. **Research data**. 2015. Disponível em: https://dictionary.casrai.org/Research_data. Acesso em: 25 set. 2018.

THE OPEN DEFINITION. **The open definition**. [2019?]. Disponível em: <https://opendefinition.org/>. Acesso em: 28 ago. 2019.

TOMOYOSE, K. **Linked Data e Classificação**: relações dialógicas com a Classificação Decimal Universal (CDU). 2018. 70f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia e Ciência da Informação) - Universidade Federal de São Carlos, São Carlos, 2018.

TORINO, E.; VIDOTTI, S. A. B. G.; SANT'ANA, R. C. G. Ciclo de vida de dados no processo de publicação e acesso à produção científica. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 20., 2019, Florianópolis. **Anais** [...]. Florianópolis: UFSC, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/576/612>. Acesso em: 14 out. 2020.

UKOLN. **The Reserach360 project**. 2013. Disponível em: <http://www.ukoln.ac.uk/projects/research360/>. Acesso em: 29 jun. 2020.

UNITED STATES DEPARTMENT OF AGRICULTURE. **About the Ag Data Commons**. [2020?]. Disponível em: <https://data.nal.usda.gov/about-ag-data-commons>. Acesso em: 22 dez. 2020.

UNITED STATES DEPARTMENT OF AGRICULTURE. **[Registro do dataset 'USDA's Expanded Flavonoid Database for the Assessment of Dietary Intakes – September 2014']**. 2019. Disponível em: <https://data.nal.usda.gov/dataset/usdas-expanded-flavonoid-database-assessment-dietary-intakes-september-2014>. Acesso em: 22 dez. 2020.

VAN DEN EYDEN, V. *et al.* **Managing and sharing data**. 3rd ed. Colchester, UK: UK Data Archive, 2011. Disponível em: <https://ukdataservice.ac.uk/media/622417/managingsharing.pdf>. Acesso em: 16 set. 2019.

W3C DATASET EXCHANGE WORKING GROUP. **[DCAT examples]**. c2020. Disponível em: <https://github.com/w3c/dxwg/tree/gh-pages/dcat/examples>. Acesso em: 14 jul. 2020.

WEIBEL, S. *et al.* **Dublin Core metadata for resource discovery**. 1998. Disponível em: <https://tools.ietf.org/html/rfc2413>. Acesso em: 18 mar. 2020.

WERSIG, G. Information science: the study of postmodern knowledge usage. **Information processing & Management**, v. 29, n. 2, p. 229-239, 1993. DOI: [https://doi.org/10.1016/0306-4573\(93\)90006-Y](https://doi.org/10.1016/0306-4573(93)90006-Y).

WILKINSON, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, 2016. DOI: <https://doi.org/10.1038/sdata.2016.18>. Disponível em: <https://www.nature.com/articles/sdata201618.pdf>. Acesso em: 16 jan. 2020.

WORLD WIDE WEB CONSORTIUM. **Describing linked datasets with the VOID vocabulary**. 2011. Disponível em: <http://www.w3.org/TR/2011/NOTE-void-20110303/>. Acesso em: 7 ago. 2018.

WORLD WIDE WEB CONSORTIUM. **Introduction to SKOS**. 2012a. Disponível em: <https://www.w3.org/2004/02/skos/intro>. Acesso em: 29 abr. 2020.

WORLD WIDE WEB CONSORTIUM. **Linked Data**. C2015a. Disponível em: <https://www.w3.org/standards/semanticweb/data>. Acesso em: 28 ago. 2019.

WORLD WIDE WEB CONSORTIUM. **OWL 2 web ontology language document overview (second edition)**. 2012b. Disponível em: <http://www.w3.org/TR/2012/REC-owl2-overview-20121211/>. Acesso em: 27 maio 2020.

WORLD WIDE WEB CONSORTIUM. **Schema**. 2015^a. Disponível em: <https://www.w3.org/standards/xml/schema>. Acesso em: 9 ago. 2018.

WORLD WIDE WEB CONSORTIUM. **Semantic Web**. C2015b. Disponível em: <https://www.w3.org/standards/semanticweb/>. Acesso em: 28 ago. 2019.

WORLD WIDE WEB CONSORTIUM. **Vocabularies**. C2015c. Disponível em: <https://www.w3.org/standards/semanticweb/ontology>. Acesso em: 10 set. 2019.

WORLD WIDE WEB CONSORTIUM. **XML essentials**. 2015b. Disponível em: <https://www.w3.org/standards/xml/core>. Acesso em: 4 jul. 2018.

WOODLEY, M. S.; CLEMENT, G.; WINN, P. **Using Dublin Core**. 2005. Disponível em: <http://www.dublincore.org/documents/2005/11/07/usageguide/glossary/>. Acesso em: 13 mar. 2020.

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

ZIEGLER, M. F. **FAPESP lança Rede de Repositórios de Dados Científicos do Estado de São Paulo**. 2019. Disponível em: <https://agencia.fapesp.br/fapesp-lanca-rede-de-repositorios-de-dados-cientificos-do-estado-de-sao-paulo/32251/>. Acesso em: 07 out. 2020.