



**PROJETO DE GRADUAÇÃO**

**REQUISITOS DE INTEROPERABILIDADE E QUALIDADE  
DE DADOS PARA A CRIAÇÃO DE RECURSOS  
EDUCACIONAIS ABERTOS CONECTADOS**

**Pamella Oliveira de Souza**

**12/0131617**

**Brasília, Maio de 2021**

**UNIVERSIDADE DE BRASÍLIA**

**FACULDADE DE TECNOLOGIA  
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO**

UNIVERSIDADE DE BRASÍLIA  
Faculdade de Tecnologia

PROJETO DE GRADUAÇÃO

**REQUISITOS DE INTEROPERABILIDADE E QUALIDADE  
DE DADOS PARA A CRIAÇÃO DE RECURSOS  
EDUCACIONAIS ABERTOS CONECTADOS**

**Pamella Oliveira de Souza**

**12/0131617**

*Projeto de Graduação submetido ao Departamento de Engenharia  
de Produção como requisito parcial para obtenção  
do grau de Engenheira de Produção*

Banca Examinadora

Prof. Edgard Costa Oliveira, EPR/UnB  
*Orientador*

\_\_\_\_\_

Prof. Simone Borges Simão Monteiro, EPR/UnB  
*Examinador interno*

\_\_\_\_\_

Prof. Edison Ishikawa, CIC/UnB  
*Examinador externo*

\_\_\_\_\_

## FICHA CATALOGRÁFICA

SOUZA, PAMELLA OLIVEIRA DE  
REQUISITOS DE INTEROPERABILIDADE E QUALIDADE DE DADOS PARA A CRIAÇÃO DE RECURSOS EDUCACIONAIS ABERTOS CONECTADOS [Distrito Federal] 2021.  
xvi, 86 p., 210 x 297 mm (EPR/FT/UnB, Engenharia, Engenharia de Produção, 2021).  
Projeto de Graduação - Universidade de Brasília, Faculdade de Tecnologia.  
Departamento de Engenharia de Produção

1. Recursos Educacionais Abertos	2. Dados Conectados
3. Interoperabilidade	4. Qualidade de Dados
I. EPR/FT/UnB	II. Título (série)

## REFERÊNCIA BIBLIOGRÁFICA

SOUZA, P. O. (2021). *REQUISITOS DE INTEROPERABILIDADE E QUALIDADE DE DADOS PARA A CRIAÇÃO DE RECURSOS EDUCACIONAIS ABERTOS CONECTADOS*. Projeto de Graduação, Departamento de Engenharia de Produção, Universidade de Brasília, Brasília, DF, 86 p.

## CESSÃO DE DIREITOS

AUTOR: Pamella Oliveira de Souza

TÍTULO: REQUISITOS DE INTEROPERABILIDADE E QUALIDADE DE DADOS PARA A CRIAÇÃO DE RECURSOS EDUCACIONAIS ABERTOS CONECTADOS.

GRAU: Engenharia de Produção ANO: 2021

É concedida à Universidade de Brasília permissão para reproduzir cópias desta Projeto de Graduação e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Os autores reservam outros direitos de publicação e nenhuma parte dessa Projeto de Graduação pode ser reproduzida sem autorização por escrito dos autores.

---

Pamella Oliveira de Souza

Depto. de Engenharia de Produção (EPR) - FT

Universidade de Brasília (UnB)

Campus Darcy Ribeiro

CEP 70919-970 - Brasília - DF - Brasil

## DEDICATÓRIA

*Dedico este trabalho a minha mãe, Maria Socorro, e a minha irmã, Paloma, que sempre me apoiaram incondicionalmente.*

## **AGRADECIMENTOS**

A Deus, por ser minha base e fortaleza.

A minha família, em especial a minha mãe, Maria Socorro, e a minha irmã, Paloma, por todo o amor e apoio.

Aos meus amigos Luiz Fernando, Erica Oliveira, Diego Neves e Aletéia Melo por sempre estarem do meu lado.

Ao meu orientador, professor Edgard, pelo empenho nessa jornada.

Ao mestrando Everaldo, que foi um grande aliado no desenvolvimento deste projeto.

A UnB, onde tanto aprendi como profissional e como pessoa.

---

## RESUMO

Os Recursos Educacionais Abertos surgiram como uma proposta para impulsionar uma educação inclusiva e de qualidade. O movimento incentiva que conteúdos usados para aprendizagem, ensino e pesquisa possam ser acessados, adaptados e distribuídos livremente. Desde que foi criado, em 2002, a iniciativa vem ganhando entusiastas dentro da sociedade civil, empresas e governos. Entretanto, apesar da quantidade de conteúdos produzidos, não há um aproveitamento eficaz destes dados devido, principalmente, à falta de padronização dos recursos públicos na Web. A Web semântica se apresenta como uma solução para a problemática ao buscar conectar os dados através de padrões e protocolos que permitem que, não apenas humanos possam processar e mesclar os dados, mas também as máquinas. Dessa forma, a padronização de dados na web é essencial para a interoperabilidade e qualidade deste. Este trabalho tem por objetivo levantar requisitos que corroboram para a geração de Recursos Educacionais Abertos Conectados com qualidade e que sejam interoperáveis, de forma a otimizar o potencial de uso destes recursos. Para tal, foram pesquisadas boas práticas e padrões de interoperabilidade e qualidade para publicação de dados na web, elicitados requisitos para a geração de dados de acordo com as recomendações do W3C (World Wide Web Consortium) e, por fim, foi proposta a representação semântica para o registro de recursos educacionais tendo como base as necessidades do Projeto PUMA (Plataforma Unificada de Metodologia Ativa). Essa representação permitiu aplicar os requisitos listados e apoiará o projeto na criação de um banco de dados robusto e interoperável com outros conjuntos de dados. Para a construção da proposta, as informações foram obtidas através de levantamento bibliográficos e entrevistas. A pesquisa foi classificada como de Natureza Aplicada, Forma de Abordagem Qualitativa, Objetivos Exploratórios e Procedimento técnico Pesquisa-Ação.

**Palavras-chaves:** Recursos Educacionais Abertos; Dados Conectados; Qualidade de Dados; Interoperabilidade.

---

## ABSTRACT

Open Educational Resources emerged as a proposal to promote inclusive and quality education. The movement encourages that learning, teaching, and research content to be able to be accessed, adapted and distributed freely. Since it was created in 2002, the initiative has been gaining enthusiasts within civil society, companies, and governments. However, despite the amount of content produced, there is no effective use of this data due, mainly, to the lack of standardization of public resources on the Web. The semantic Web presents itself as a solution to the problem when trying to connect the data through standards and protocols that allow not only humans to process and merge data, but also machines. Thus, the standardization of data on the web is essential for its interoperability and quality. This project aims to raise requirements that corroborate the generation of Connected Open Educational Resources with quality and that are interoperable, in order to optimize the potential use of these resources. To this end, good practices, interoperability, and quality standards for publishing data on the web were researched, requirements for data generation were elicited according to the recommendations of the W3C (World Wide Web Consortium) and, finally, the semantic representation was proposed for the registration of educational resources based on the needs of the PUMA Project (Unified Platform for Active Methodology). This representation allowed applying the listed requirements and will support the project in creating a robust and interoperable database with other datasets. For the construction of the proposal, the information was obtained through bibliographic surveys and interviews. The research was classified as Applied Nature, Qualitative Approach Form, Exploratory Objectives, and Action Research Technical Procedure.

**Key-words:** Open Educational Resources; Linked Data; Data Quality; Interoperability.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
1.1	CONTEXTUALIZAÇÃO	1
1.2	DEFINIÇÃO DA PROBLEMÁTICA	2
1.3	JUSTIFICATIVA DO TEMA	3
1.4	OBJETIVOS	4
1.4.1	OBJETIVO GERAL	4
1.4.2	OBJETIVOS ESPECÍFICOS	4
1.5	PROCEDIMENTOS METODOLÓGICOS	4
1.5.1	ESTRUTURAÇÃO DA PESQUISA	6
1.6	ESTRUTURA DOS CAPÍTULOS	7
<b>2</b>	<b>REVISÃO SISTEMÁTICA DA LITERATURA</b>	<b>8</b>
2.1	DESCRIÇÃO DO MODELO TEMAC	8
2.1.1	ETAPA 1: PREPARAÇÃO DA PESQUISA	9
2.1.2	ETAPA 2: APRESENTAÇÃO E INTERRELAÇÃO DOS DADOS	10
2.1.3	ETAPA 3: DETALHAMENTO, MODELO INTEGRADOR E VALIDAÇÃO POR EVIDÊNCIAS	11
2.2	APLICAÇÃO DA METODOLOGIA	11
2.2.1	ETAPA 1: PREPARAÇÃO DA PESQUISA	11
2.2.2	ETAPA 2: APRESENTAÇÃO E INTERRELAÇÃO DOS DADOS	12
2.2.3	ETAPA 3: DETALHAMENTO, MODELO INTEGRADOR E VALIDAÇÃO POR EVIDÊNCIAS	16
<b>3</b>	<b>REFERENCIAL TEÓRICO</b>	<b>21</b>
3.1	DADOS CONECTADOS	21
3.1.1	DADOS ABERTOS	22
3.1.2	WEB SEMÂNTICA	28
3.1.3	ESTRUTURAÇÃO DE DADOS CONECTADOS	31
3.1.4	MELHORES PRÁTICAS PARA PUBLICAÇÃO DE DADOS CONECTADOS	33
3.2	INTEROPERABILIDADE DE DADOS	34
3.3	RECURSOS EDUCACIONAIS ABERTOS	35
3.4	QUALIDADE DE DADOS	37
<b>4</b>	<b>BOAS PRÁTICAS E PADRÕES DE INTEROPERABILIDADE E QUALIDADE DE DADOS</b>	<b>40</b>
4.1	BOAS PRÁTICAS PARA DADOS NA WEB	40
4.2	VOCABULÁRIOS PADRONIZADOS	49
4.2.1	DUBLIN CORE METADATA INITIATIVE (DC)	50
4.2.2	FRIEND OF A FRIEND VOCABULARY (FOAF)	51
4.2.3	SIMPLE KNOWLEDGE ORGANIZATION SYSTEM (SKOS)	52



4.2.4	THE PROV ONTOLOGY (PROV-O).....	53
4.2.5	DATA CATALOG VOCABULARY (DCAT) .....	55
4.3	REPRESENTAÇÕES SINTÁTICAS PARA O MODELO RDF 1.1.....	58
4.3.1	RDF/XML .....	58
4.3.2	RDFa .....	59
4.3.3	JSON-LD .....	60
4.3.4	N-TRIPLES .....	61
4.3.5	N-QUADS .....	61
4.3.6	TURTLE .....	62
4.3.7	TRIG .....	62
<b>5</b>	<b>ELICITAÇÃO E APLICAÇÃO DE REQUISITOS PARA GERAÇÃO DE RECURSOS EDUCACIONAIS ABERTOS CONECTADOS .....</b>	<b>64</b>
5.1	ELICITAÇÃO DE REQUISITOS .....	64
5.1.1	METADADOS .....	66
5.1.2	FORMATOS DE DADOS .....	67
5.1.3	USO DE VOCABULÁRIOS PADRÕES .....	67
5.1.4	LICENÇA DE DADOS .....	68
5.1.5	IDENTIFICADORES DE DADOS.....	68
5.1.6	COLETA E COMPARTILHAMENTO DE FEEDBACKS.....	68
5.2	PROPOSTA DE REPRESENTAÇÃO SEMÂNTICA PARA REGISTRO DE RECURSOS EDUCACIONAIS .....	69
5.2.1	PLATAFORMA UNIFICADA DE METODOLOGIA ATIVA (PUMA) .....	69
5.2.2	PROPOSTA DE REPRESENTAÇÃO DE REGISTRO DE PROJETO .....	71
<b>6</b>	<b>CONSIDERAÇÕES FINAIS, LIMITAÇÕES E FUTURAS LINHAS DE PESQUISA. ....</b>	<b>75</b>
6.1	LIMITAÇÕES DA PESQUISA .....	75
6.2	FUTURAS LINHAS DE PESQUISA .....	76
	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>77</b>
	<b>APÊNDICES .....</b>	<b>85</b>
<b>I</b>	<b>PUBLICAÇÕES E CITAÇÕES SOBRE O TEMA RECURSOS EDUCACIONAIS ABERTOS .....</b>	<b>86</b>

# LISTA DE FIGURAS

1.1	Diagrama das etapas de realização da pesquisa. ....	7
2.1	Representação das etapas do Modelo TEMAC.....	9
2.2	Publicações ano a ano sobre os termos Recursos Educacionais abertos e Dados Conectados. ....	13
2.3	<i>Word Cloud</i> com as palavras-chaves dos artigos.....	16
2.4	Co-citações encontradas no artigos.....	17
2.5	<i>Acoplamento bibliográfico</i> . ....	18
2.6	Palavras mais frequentes.....	19
3.1	Ciclo de Vida da Informação. ....	22
3.2	Classificação dos países avaliados no <i>Open Data Barometer - Leaders Edition</i> . ....	25
3.3	Sistema de 5 estrelas para dados abertos. ....	28
3.4	Camadas da Web Semântica Revisada. ....	29
3.5	Gráfico RDF relacionando recursos.....	31
3.6	Exemplo de tripla RDF.....	32
3.7	Múltiplas triplas RDF.....	32
3.8	Dimensões da qualidade de dados. ....	39
4.1	Representação das interações entre os elementos de dados no DWBP. ....	40
4.2	Elementos básicos do vocabulário PROV-O: entidade, atividade e agente. ....	53
4.3	Diagrama de Classes de DCAT.....	57
4.4	Formatos de serialização das versões RDF1 e RDF1.1.....	58
5.1	Princípios e melhores praticas para Dados Abertos, Dados Conectados e Dados na Web. ....	65
5.2	Exemplos de metadados descritivos. ....	66
5.3	Esquema Geral para Projeto âncoras (y, z) do curso de Engenharia de Produção da UnB. ....	70
5.4	Página Web legível para humanos. ....	71
5.5	Programação RDFa para registros de Recursos Educacionais Abertos.....	73
5.6	Formato de registro de feedback e licença do item.....	74
I.1	Número de publicações ano a ano para Recursos Educacionais Abertos. ....	86
I.2	Número de citações ano a ano para Recursos Educacionais Abertos. ....	86

# LISTA DE TABELAS

1.1	Classificação de pesquisas científicas.....	5
1.2	Classificação da pesquisa desenvolvida. ....	6
2.1	Definições das revisões quantitativa, integrativa, sistemática e meta-análises. ....	9
2.2	Revistas mais relevantes. ....	12
2.3	Publicações por países.....	13
2.4	Organizações que mais publicaram .....	14
2.5	Áreas que mais publicam. ....	14
2.6	Artigos mais citados. ....	15
2.7	Enfoque dos artigos mais citados. ....	20
3.1	Marcos da abertura de Dados Governamentais.....	25
4.1	Benefícios das Boas Práticas.....	42
4.2	Correlação entre dimensões da qualidade de dados Boas práticas para Dados.....	44
4.3	Lista de vocabulários com seus respectivos prefixos e URIs.....	49
4.4	Namespaces do Vocabulário Dublin Core. ....	51
4.5	Classes e Propriedades do Vocabulário FOAF. ....	52
4.6	Classificação, classes e propriedades do vocabulário PROV-O .....	54
4.7	Classes e propriedades do DCAT. ....	56
4.8	Dados utilizados para descrição do Código RDF/XML.....	59
5.1	Requisitos para a geração de Recursos Educacionais Abertos Conectados. ....	65
5.2	Definição e descrição dos metadados do Projeto PUMA. ....	72
5.3	Termos e vocabulários dos metadados.....	73

# LISTA DE SÍMBOLOS

## Siglas

CSV	<i>Comma-Separated Values</i>
DC	<i>Dublin Core</i>
DCAT	<i>Data Catalog Vocabulary</i>
ePING	Padrões de Interoperabilidade de Governo Eletrônico
FOAF	<i>Friend of a Friend Vocabulary</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
INDA	Infraestrutura Nacional de Dados Abertos
JSON	<i>JavaScript Object Notation</i>
LAI	Lei de Acesso à Informação
MIT	<i>Massachusetts Institute of Technology</i>
ODC	<i>Open Data Charter</i>
ODS	<i>Open Document Spreadsheet</i>
ONU	Organização das Nações Unidas
PBL	<i>Project Based Learning</i>
PROV-O	<i>PROV Ontology</i>
PSPs	Projetos de Sistemas de Produção
PUMA	Plataforma Unificada de Metodologia Ativa
RDF	<i>Resource Description Framework</i>
REA	Recursos Educacionais Abertos
SKOS	<i>Simple Knowledge Organization System</i>
SLTI	Secretaria de Logística e Tecnologia da Informação
TCU	Tribunal de Contas da União
TEMAC	Teoria do Enfoque Meta Analítico Consolidado
UnB	Universidade de Brasília
UNESCO	<i>United Nations Educational, Scientific and Cultural Organization</i>
URI	<i>Uniform Resource Identifier</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>Extensible Markup Language</i>

# 1 INTRODUÇÃO

O presente trabalho tem como objetivo a especificação de requisitos para geração de Recursos Educacionais Abertos com qualidade e interoperabilidade de dados. Este Capítulo apresenta a contextualização do assunto, definição da problemática, justificativa para desenvolvimento da pesquisa, objetivos gerais e específicos, procedimentos metodológicos, as etapas de desenvolvimento do projeto e a estrutura em que o trabalho foi organizado.

## 1.1 CONTEXTUALIZAÇÃO

A iniciativa Recursos Educacionais Abertos (REA) surgiu em 2002 durante uma convenção promovida pela UNESCO em parceria com o MIT (*Massachusetts Institute of Technology*). O termo se trata de uma filosofia que busca promover o conhecimento livre de barreiras, podendo os dados educacionais serem usados e reutilizados para a aprendizagem, o ensino e a pesquisa através de licenças abertas ou sob domínio público. De acordo com a (UNESCO 2012), “os Recursos Educacionais Abertos fornecem uma oportunidade estratégica para melhorar a qualidade da educação, bem como facilitar o diálogo sobre políticas, compartilhamento de conhecimento e capacidade de construção”.

Apesar dos esforços empenhados por governos, instituições e indivíduos interessados em promover uma educação mais igualitária e inclusiva por meio dos REA (Amiel, Gonsales e Sebriam 2018), de nada adianta se estes recursos não puderem ser utilizados. Segundo (Bittencourt et al. 2008), a quantidade de informações produzidas pela humanidade torna praticamente impossível o processamento e aproveitamento dessas, sendo necessário que agentes de *software* trabalhem para processar esses conjuntos de dados de forma eficaz. Nesse contexto, a Web Semântica se apresenta como uma solução para a geração de valor de dados. Sendo considerada uma evolução da Web atual, a Web semântica busca, de forma colaborativa, organizar as informações para que elas sejam legíveis tanto para humanos quanto para máquinas (Bizer, Heath e Berners-Lee 2011). Dessa forma, os dados podem interagir dentro da Web por meio de agentes de *software*, um conceito conhecido como interoperabilidade de dados.

A Web Semântica é tida como uma visão do W3C (*World Wide Web Consortium*) para Dados Conectados (W3C 2015). Dados Conectados (do inglês, *Linked Data*) pode ser definido com um conjunto de boas práticas para a publicação e conexão de um grupo de dados estruturados (Bizer, Heath e Berners-Lee 2011). Portanto, os Dados Conectados estabelecem “princípios para a publicação e consumo dos dados e os classificam de acordo com sua disponibilidade, acesso, estruturação e conexão”(Open Knowledge Brasil 2017). O conceito foi criado por Tim Berners-Lee devido a necessidade da criação de padrões dentro da Web de dados (Isotani e Bittencourt 2015).

Corroborando com a expansão dos Dados Conectados e consequente consolidação da Web semântica, o movimento Dados Abertos, que vem ganhando força desde 2009 com iniciativas de governos, sociedade civil e entusiastas. De acordo com a Open Knowledge Foundation, “dados e conteúdos abertos podem ser usados, modificados e compartilhados livremente por qualquer pessoa para qualquer propósito”, os dados abertos devem ser completos, primários, atuais, acessíveis, processáveis por máquina, terem acesso não discriminatório, formatos não proprietários e possuir licenças livres (Portal Brasileiro de Dados Abertos 2015), possibilitando que organizações e pessoas consigam, tranquilamente, acessar informações para produzir novos conhecimentos, tomar decisões e gerar inovações, por exemplo (Bandeira et al. 2014).

Nesse contexto, Recursos Educacionais Abertos podem ser considerados verdadeiramente abertos se seguem os princípios dos Dados Abertos e, ademais, utilizam padrões que permitam aos seus dados serem processados tanto por humanos, quanto por máquinas garantindo assim, a interoperabilidade entre diferentes coleções desses recursos educacionais (Piedra et al. 2015).

Segundo (Chicaiza et al. 2014), uma das razões que levam os Recursos Educacionais Abertos a serem mal utilizados é a falta de padronização das instituições produtoras de conhecimento no momento de criar os dados, usando de “etiquetas ou esquemas informais”. O uso de dados abertos aplicados à educação é uma área relativamente nova e com potencial de uso não totalmente vislumbrado (Atenas et al. 2015, Penteado, Bittencourt e Isotani 2019, Coughlan 2020).

No que concerne à qualidade de dados, não há um consenso sobre o conceito. A qualidade é mensurada a partir de dimensões da qualidade que representam características importantes para consumidores e publicadores (Lóscio, Burle e Calegari 2017). Os pesquisadores (Valente e Fujino 2016) apontam que, dentro da área da ciência da informação, existem vários critérios para qualificar dados e informações. Esses critérios são utilizados para “definir, medir e gerenciar a qualidade e variam de acordo com as abordagens e vertentes sob as quais os estudos são realizados” (Fagundes, Macedo e Freund 2018).

## **1.2 DEFINIÇÃO DA PROBLEMÁTICA**

A educação é uma das áreas que mais se beneficia com a abertura, compartilhamento e interoperabilidade de dados, posto que o acesso livre a recursos educacionais possui um profundo impacto na geração de conhecimento. Entretanto, a falta de padronização entre os publicadores de dados dificulta a utilização eficaz dos recursos gerados (Nahhas et al. 2018).

Apesar do extremo potencial e relevância que tem o uso de Dados Abertos Conectados para a otimização do conhecimento gerado no âmbito educacional, essa área ainda é pouco explorada.

Este trabalho busca responder o seguinte questionamento: quais são os requisitos que permitem a criação de Recursos Educacionais Abertos Conectados com qualidade<sup>1</sup>, possibilitando a interoperabilidade dos dados gerados.

### 1.3 JUSTIFICATIVA DO TEMA

O trabalho é embasado na necessidade de geração de Dados Abertos Conectados com qualidade para melhor promover o aproveitamento das informações oriundas de recursos educacionais, possibilitando, dessa maneira, a interoperabilidade destes dados. (Love et al. 2016) relata as dificuldades de encontrar dados abertos apropriados para a mineração, sendo necessário que os educadores trabalhem esses dados com antecipação. Assim, é necessária a preparação dos dados de modo a otimizar a mineração e a geração de valor.

Este Projeto de Graduação justifica-se socialmente pela importância da gestão de Recursos Educacionais Abertos Conectados para a sociedade como um todo, trazendo como benefícios uma maior transparência das atividades educacionais geradas dentro da universidade, a possibilidade de inovação de trabalhos científicos ao permitir conhecer os trabalhos já realizados, assim com as linhas de pesquisas mais atuais, o incentivo ao aprimoramento de dados ao zelar pela qualidade dos dados criados e, em um contexto de extensão universitária, a promoção da criação de novos negócios por meio da aplicação do conhecimento científico produzido. Ademais, o trabalho visa incentivar a educação de qualidade, meta número 4 da ONU (ONU Brasil 2012). (Guterres 2020), atual secretário geral da ONU, afirma que “a educação é a chave para o desenvolvimento pessoal e o futuro das sociedades. Desbloqueia oportunidades e reduz desigualdades. É o alicerce das sociedades informadas e tolerantes e o principal impulsionador do desenvolvimento sustentável”.

O trabalho justifica-se cientificamente pela relevância do tema para o meio acadêmico. Uma pesquisa realizada no *ISI Web of Science-WoS*, considerando os últimos 15 anos, sobre o termo "*Open Educational Resources*" retornou um total de 1.350 publicações e 5.926 citações. As Figuras I.1 e I.2, em Apêndice, apresentam esses dados ano a ano. É possível notar um crescimento anual das publicações e citações, o que confirma a importância do tema.

Em relação ao Curso de Engenharia de Produção, a justificativa está pautada na aplicação dos conhecimentos adquiridos durante o curso, principalmente nas áreas de Gestão do Conhecimento Organizacional, Educação em Engenharia de Produção e Gestão da Qualidade. Além disso, o trabalho beneficia professores e alunos do curso ao propiciar ferramentas de geração de Recursos Educacionais Abertos Conectados com qualidade e, em um momento futuro, a possibilidade de geração desses dados à comunidade acadêmica.

---

<sup>1</sup>As dimensões de qualidade de dados consideradas neste projeto são do modelo de (Zaveri et al. 2016), adaptado de (Wang e Strong 1996), que será explicitado na seção 3.4.

## **1.4 OBJETIVOS**

### **1.4.1 Objetivo Geral**

Levantar requisitos para a criação de dados em Recursos Educacionais Abertos Conectados, visando a interoperabilidade e a qualidade da informação em projetos educacionais.

### **1.4.2 Objetivos Específicos**

Para o cumprimento dos objetivos gerais, estabelece-se como objetivos específicos:

1. Pesquisar padrões e boas práticas de interoperabilidade e qualidade de dados para documentos semanticamente representáveis;
2. Elicitar requisitos de interoperabilidade e qualidade para recursos educacionais;
3. Propor uma representação semântica para registro de projetos educacionais;
4. Aplicar a representação proposta em uma plataforma educacional.

## **1.5 PROCEDIMENTOS METODOLÓGICOS**

O conhecimento científico tem relação com a lógica e o pensamento analítico, compreendendo informações e fatos que são comprovados com base em uma série de experimentações e testes. São características do conhecimento científico: pesquisas realizadas com fatos, elaboração de proposições ou hipóteses, estudos sistemáticos e verificáveis. Além disso, esse conhecimento não é definitivo, podendo ser atualizado por outros estudos (Pereira et al. 2018). Segundo (Martins, Mello e Turrioni 2013), a ciência usa de investigação metódica da realidade para entender as leis que regem a natureza, assim como seus fenômenos e processos e dessa forma, conseguir beneficiar a humanidade. Através de reflexão e entendimento dos fenômenos, é possível gerar novos conhecimentos.

Segundo (Gil 2017), pesquisa é definida como “o procedimento racional e sistemático que tem como objetivo proporcionar respostas aos problemas que são propostos”, buscando resolver problemas ainda sem solução ou quando, apesar da problemática possuir boa qualidade de informações, essas não se aplicam as questões específicas trabalhadas pelo pesquisador. Os pesquisadores (Silva e Menezes 2005) categorizam as pesquisas de acordo com sua natureza, forma de abordagem, objetivos e procedimentos técnicos, conforme Tabela 1.1.



Tabela 1.1: Classificação de pesquisas científicas.

CATEGORIA	CLASSIFICAÇÃO
Natureza	Básica
	Aplicada
Forma de abordagem	Quantitativa
	Qualitativa
Objetivos	Exploratória
	Descritiva
	Explicativa
Procedimentos técnicos	Pesquisa Bibliográfica
	Pesquisa Documental
	Pesquisa Experimental
	Levantamento
	Estudo de caso
	Pesquisa Expost-Facto
	Pesquisa-Ação
	Pesquisa Participante

Fonte: (Silva e Menezes 2005).

A presente pesquisa foi classificada como de Natureza Aplicada posto que objetiva gerar conhecimento prático para a resolução da problemática apresentada. Sua forma de abordagem é Qualitativa considerando que não se utiliza de técnicas e métodos estáticos e está pautada na observação de fenômenos e seus significados.

Em relação aos objetivos, a pesquisa é Exploratória ao propiciar uma maior familiaridade com o tema apresentado através de levantamento bibliográfico e entrevistas com um representante do Projeto PUMA. O Projeto PUMA é pautado no desenvolvimento de uma plataforma que busca integrar os resultados da aplicação da metodologia ativa de ensino por meio de projetos (PBL). O representante do projeto foi o mestrando Everaldo Júnior Silva, que está desenvolvendo uma tese denominada *Tratamento de riscos relacionados à alocação de projetos: proposição de módulo para uma plataforma de aprendizagem ativa*. Essa tese tem como objetivo propor uma ferramenta computacional que trate os riscos da alocação incorreta de *stakeholders* dentro das disciplinas de Projetos de Sistemas de Produção (PSPs) e Projetos de Graduação (PGs).

As entrevistas proporcionaram relatos sobre experiências práticas em relação ao tema pesquisado. Ao todo foram realizadas 5 reuniões com tempo aproximado de 1h cada para entender melhor as necessidades do Projeto. As entrevistas foram não-estruturadas. De acordo com (Aguiar e Medeiros 2009), a entrevista não estruturada está baseada na flexibilidade e busca de significado advinda do entrevistado. Diferente de entrevistas estruturadas e semi-estruturadas, não há um roteiro pré-definido. O entrevistado é encorajado a falar sobre o tema de pesquisa de forma livre, que permite a ele expressar suas considerações de acordo com suas próprias referências (May 2004).

Por fim, o seu procedimento técnico é uma Pesquisa-Ação. Esse procedimento visa realizar uma “estreita associação com uma ação ou com a resolução de um problema coletivo”, onde os pesquisadores e representantes do problema estão envolvidos de forma colaborativa (Silva e Menezes 2005). A Tabela 1.2

sintetiza a classificação da pesquisa.

Tabela 1.2: Classificação da pesquisa desenvolvida.

CATEGORIA	CLASSIFICAÇÃO
Natureza	Aplicada
Forma de abordagem	Qualitativa
Objetivos	Exploratória
Procedimento técnico	Pesquisa-Ação

Fonte: Autoria própria.

### 1.5.1 Estruturação da pesquisa

Com o propósito de alcançar os objetivos determinados, a pesquisa foi desenvolvida em seis etapas, conforme descrito a seguir:

- **1ª Etapa - Entendimento do Contexto:** O problema de pesquisa surgiu após a observação da dificuldade de geração de dados com qualidade dentro de projetos educacionais. No primeiro momento procurou-se entender a problemática através de pesquisas iniciais sobre o tema e entrevistas com um representante do Projeto PUMA;
- **2ª Etapa - Revisão Sistemática da Literatura:** Em um segundo momento, realizou-se uma revisão sistemática da literatura. Ao utilizar o modelo Teoria do Enfoque Meta Analítico Consolidado (TEMAC), foi possível encontrar fontes que estivessem mais alinhadas aos objetivos da pesquisa, além de permitir dispor de artigos atualizados e reconhecidos pelo meio acadêmico;
- **3ª Etapa - Levantamento Bibliográfico:** Após a revisão sistemática da literatura, foi realizada o Levantamento Bibliográfico, onde investigou-se de forma mais profunda sobre os temas base para a pesquisa: Dados Conectados, Interoperabilidade de Dados, Recursos Educacionais Abertos e Qualidade de Dados. Ademais dos arquivos encontrados na revisão sistemática da Literatura, foram utilizadas as citações dos artigos e principalmente documentação advinda do W3C.
- **4ª Etapa - Determinação de Padrões e boas práticas para geração de dados:** A próxima etapa foi pautada pela pesquisa e especificação de boas práticas e padrões necessários para interoperabilidade e qualidade de dados de acordo com as normas estabelecidas pelo W3C;
- **5ª Etapa - Definição de Requisitos:** Na seguinte etapa, definiu-se requisitos para a geração de Recursos Educacionais Abertos Conectados por meio da análise dos Princípios para Dados Abertos, Melhores Práticas para Publicação de Dados Conectados e as Boas Práticas para Dados na Web;
- **6ª Etapa - Aplicação da proposta:** Por fim, desenvolveu-se e aplicou-se uma proposta para representação semântica de registro de recursos educacionais. Este modelo levou em consideração as necessidades do Projeto PUMA em relação ao cadastro dos resultados obtidos nas disciplinas que aplicam o PBL.

A Figura 1.1 mostra a representação das etapas do desenvolvimento da pesquisa.

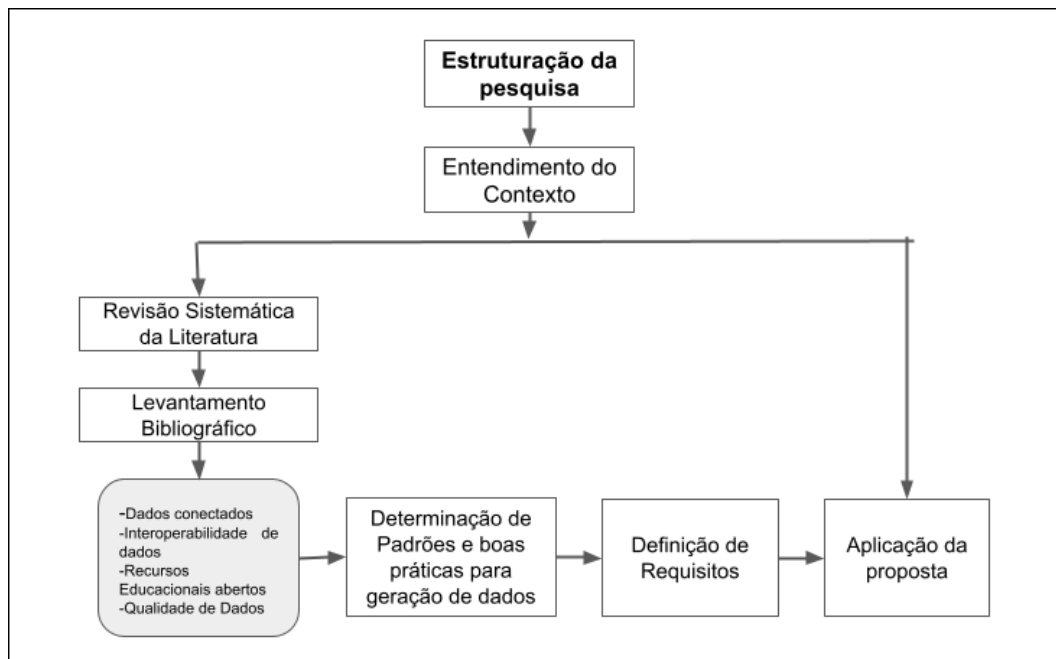


Figura 1.1: Diagrama das etapas de realização da pesquisa.

Fonte: Autoria própria.

## 1.6 ESTRUTURA DOS CAPÍTULOS

Este Projeto de Graduação estrutura-se da seguinte maneira: além da Introdução (Capítulo 1), há mais seis Capítulos. O Capítulo 2, *Revisão Sistemática da Literatura*, apresenta uma revisão estruturada sobre o tema abordado. O Capítulo 3, *Referencial Teórico*, mostra o embasamento teórico utilizado para a elaboração do projeto.

A seguir, os capítulos 4 e 5 trazem os resultados da pesquisa, de acordo com os objetivos listados na subseção 1.4.2. Sendo assim, o Capítulo 4 mostra as *Boas Práticas e Padrões de Interoperabilidade e Qualidade de Dados* e o Capítulo 5 apresenta *Elicitação e aplicação de requisitos para Geração de recursos educacionais abertos Conectados*. Por fim, o Capítulo 6, *Considerações Finais, Limitações e Futuras Linhas de Pesquisa*, contém as principais conclusões do trabalho, limitações encontradas durante a execução do projeto e sugestões para trabalhos futuros.

## 2 REVISÃO SISTEMÁTICA DA LITERATURA

A quantidade de informações produzidas pela humanidade dificulta o acesso a materiais de qualidade. Com o propósito de encontrar os artigos mais relevantes acerca do tema abordado, foi realizada uma Revisão Sistemática da Literatura utilizando o método TEMAC (Teoria do Enfoque Meta Analítico Consolidado).

Este Capítulo apresenta a descrição do TEMAC, assim como a aplicação das etapas desse método.

### 2.1 DESCRIÇÃO DO MODELO TEMAC

Quando se começa a trabalhar com uma problemática, há a possibilidade de não encontrar muitas informações disponíveis sobre o tema. Nesses casos, costuma-se realizar pesquisas com abordagem exploratória. Entretanto, a grande maioria dos temas já foram estudados por outros pesquisadores e por isso é necessário fazer um levantamento sobre o estado do problema pesquisado (Mariano e Rocha 2017). Desse modo, antes de iniciar a pesquisa, o pesquisador deve entender o estado da arte da problemática para que seu trabalho não seja uma repetição dos demais e esse possa agregar valor ao tema proposto (Mariano, Cruz e Gaitán 2011). Para a realização da revisão do estado da arte do tema pesquisado, utilizou-se o modelo de Teoria do Enfoque Meta Analítico Consolidado (TEMAC).

O TEMAC tem suas origens nos estudos sobre enfoque Meta-analítico proposto, principalmente, pelos autores (Arenas, García e Espasandin 2001), (Cruz e Correa 2004) e (Abramo e D'Angelo 2011). Esse tipo de revisão da literatura busca mais assertividade e objetividade nas pesquisas por meio da utilização de critérios de impacto na busca de referências bibliográficas, trazendo, dessa forma, maior qualidade às referências utilizadas durante o trabalho científico (Mariano e Rocha 2017).

A informação é tida como o maior bem da sociedade contemporânea, contudo a quantidade de informações é considerada um enorme problema. Segundo (Delfino, Neto e Sousa 2019), “a liberdade na comunicação, possível através do uso da internet e das redes sociais, associadas à ausência de hierarquia na produção e disseminação na internet potencializa o surgimento de uma inundação de informações que dificultam a capacidade de análise e crítica por parte dos seus usuários”. Desse modo, além da grande quantidade de informações, o advento da internet permitiu o compartilhamento de informações em tempo real, sendo que barreiras geográficas deixam de ser um empecilho para o fluxo de informação. Consequentemente, reconhecer pesquisas com qualidade tem se apresentando como uma tarefa cada vez mais difícil e significativa. Nesse contexto, a revisão sistemática da literatura vem como uma solução, posto que essa proporciona pesquisas com rigor metodológico que utilizam índices bibliométricos e, quando necessário, análises estatísticas (Albrecht et al. 2017).

O enfoque Enfoque Meta-analítico reúne características das revisões qualitativa, integrativa e sistemática, podendo utilizar, em investigações mais aprofundadas, a meta-análise. A Tabela 2.1 apresenta as definições das revisões citadas.

Tabela 2.1: Definições das revisões quantitativa, integrativa, sistemática e meta-análises.

TIPO DE REVISÃO	DEFINIÇÃO
Revisão qualitativa	Caracterizada por sintetizar descobertas de estudos qualitativos. É uma re-criação do meta-análises aplicado a dados qualitativos.
Revisão Integrativa	Utiliza estudos integradores de conceitos, métodos e opiniões para agrupar par, objetivar e trazer novas perspectivas sobre um tema.
Revisão Sistemática	Emprega o planejamento da pesquisa por meio de passos que possibilitam diminuir o tendência da pesquisa ao combinar os autores mais relevantes, obtendo alto grau de rigorosidade.
Meta-análises	Possibilita a integração de estudos de primários por intermédio de técnicas estatísticas, melhorando a eficácia da pesquisa.

Fonte: (Mariano e Rocha 2017). Adaptado.

O TEMAC, proposto por (Mariano e Rocha 2017), unificou contribuições do uso do enfoque meta-analítico e garantiu atributos significativos tais como precisão, robustez, validade, funcionalidade, tempo e custos propostos por (Abramo e D'Angelo 2011). Esses atributos foram alcançados por meio da ampliação da abrangência de atuação da pesquisa, possibilitando ao estudioso inserir quantos bases de dados achar conveniente para a sua pesquisa, o que permite análises holísticas do conteúdo. Também foram inseridas maiores quantidades de índices bibliométricos. O método permite, ainda, realizar interrelação entre países, autores, segmentos e resultados, facilitando as interpretações das descobertas de acordo com o segmento pesquisado. Entretanto, as maiores contribuições do TEMAC estão associadas ao tempo e ao custo: o TEMAC foi desenhado com etapas claras e com apoio de teorias bibliométricas robustas. Além disso, foram introduzidas ferramentas tecnológicas totalmente gratuitas para realização das análises.

A abordagem é dividida em 3 etapas, como pode ser visualizado na Figura 2.1.

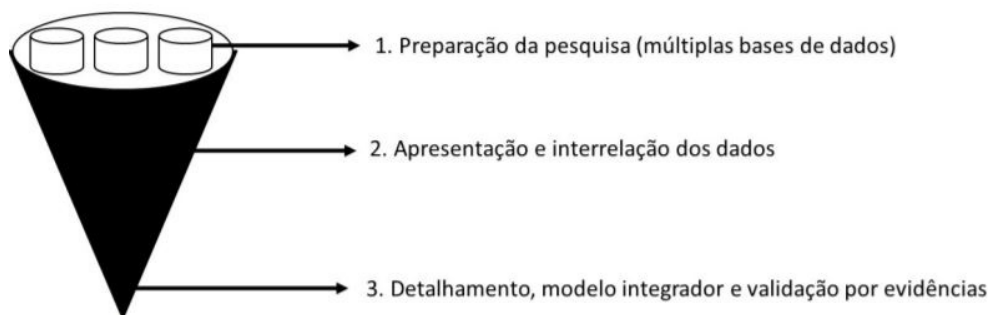


Figura 2.1: Representação das etapas do Modelo TEMAC.

Fonte: Mariano e Rocha (2017).

### 2.1.1 Etapa 1: Preparação da pesquisa

A primeira etapa do modelo consiste em responder as seguintes perguntas:

- (a) Qual o descritor, *string* ou palavra-chave da pesquisa?

- (b) Qual o campo espaço-tempo da pesquisa?
- (c) Quais as bases de dados serão utilizadas?
- (d) Quais áreas de conhecimento serão utilizadas?

Essa etapa é bastante relevante, pois a escolha das palavras de pesquisa podem alterar os resultados. Para realizar a busca preliminar, o pesquisador pode escolher um artigo base para determinar as palavras-chave que mais se adequam a sua linha de investigação. Algumas bases de dados oferecem opções de operadores de pesquisa para auxiliar nas combinações como, por exemplo, OR, AND, ( ) e “ ”.

Ao se trabalhar com diferentes bases de dados, deve-se adotar o mesmo campo espaço-tempo para todas. Tem-se adotado como campo espaço-tempo os estudos desenvolvidos em 5 e 10 anos antes a fim de se trabalhar com as pesquisas mais atualizadas, o que corrobora com a qualidade da investigação.

Ao delimitar as bases de dados que serão utilizadas na investigação, é importante justificar a escolha, posto que cada área possui plataformas online que contém uma coletânea considerável. Após realizar a busca, é necessário que o investigador faça a leitura dos títulos e, se achar necessário, use filtros segundo a área de pesquisa trabalhada. Caso a pesquisa não retorne artigos relevantes, deve-se iniciar novamente a primeira etapa, escolhendo novas palavras para executar a busca.

### **2.1.2 Etapa 2: Apresentação e interrelação dos dados**

Existem inúmeras opções de critérios que o investigador pode escolher ao realizar sua pesquisa, não obstante alguns critérios são mais comuns e se repetem nas revisões que utilizam o enfoque meta-analítico, sendo eles:

- (a) Análise das revistas mais relevantes;
- (b) Análise das revistas que mais publicam sobre o tema;
- (c) Evolução do tema ano a ano;
- (d) Países que mais publicaram;
- (e) Conferências que mais contribuíram;
- (f) Universidades que mais publicaram;
- (g) Agências que mais financiam a pesquisa;
- (h) Áreas que mais publicam;
- (i) Documentos mais citados;
- (j) Autores que mais publicaram vs. autores que mais foram citados;
- (k) Frequência de palavras chaves.

De acordo com os autores do modelo, embora a lista possa ser vista como extensa, os itens citados são importantes posto que estão atrelados a princípios ou leis bibliométricas como, por exemplo, Lei de Bradford, fator de Impacto, Lei do 80/20, Obsolescência da literatura, Teoria Epidêmica de Goffman, Lei de Lokta e Lei do Elitismo, Lei do Elitismo e citações.

Durante a coleta dos dados, o pesquisador adquire ferramentas e argumentos para inter-relacionar as informações e tornar a investigação mais robusta. É interessante que cada resultado seja analisado e comentado a fim de agregar valor ao resultado.

### **2.1.3 Etapa 3: Detalhamento, modelo integrador e validação por evidências**

A terceira etapa do método é marcada pelo aprofundamento das análises mediante índices bibliométricos que detectam os colégios invisíveis tais como Co-citação, Coupling e Co-autoria. Tais análises permitem compreender melhor o tema, assim como selecionar os principais autores por meio de comparação de resultados por fontes diferentes. Por fim, faz-se análise de co-ocorrência e frequência das palavras-chave com o propósito de determinar as linhas de pesquisas mais relevantes por meio das palavras-chave.

De acordo com (Mariano e Rocha 2017), "o pesquisador deve realizar uma catalogação pessoal dos trabalhos visando encontrar as similaridades apontadas nas análises bibliométricas anteriores (Co-citação, Coupling, coautoria, co-ocorrência e frequência de palavras-chave)."

## **2.2 APLICAÇÃO DA METODOLOGIA**

### **2.2.1 Etapa 1: Preparação da pesquisa**

De acordo com a proposta da pesquisa, as palavras selecionadas na primeira etapa foram:

1. Dados conectados;
2. Recursos Educacionais Abertos.

A base de dados escolhida para pesquisas foi a *ISI Web of Science-WoS*<sup>1</sup>. Essa plataforma foi eleita por ter uma base de dados interdisciplinar e ser um respeitável indexador de periódicos científicos (Fagundes, Macedo e Freund 2018). A base de dados possibilita ferramentas para análise de citações, referências e índice h, o que permite análises bibliométricas robustas. Outra vantagem é a cobertura de periódicos que conta com cerca 12.000 exemplares e disponibilidade de acesso desde 1945 até o presente (PORTAL DE PERIÓDICOS DA CAPES/MEC 2020).

Com o propósito de entender como os Dados Conectados têm sido utilizados com o tema principal do estudo, Recursos Educacionais Abertos, utilizou-se a combinação ("*linked data*") AND ("*Open Educational Resources*").

---

<sup>1</sup>[www.webofknowledge.com](http://www.webofknowledge.com)

As pesquisas foram realizadas entre 24 e 28 de novembro de 2020. Não foram registrados dados sobre revistas que mais publicam sobre o tema, conferências que mais contribuíram e agências que mais financiam pesquisas, conforme citado em 2.1.2, posto que a *ISI Web of Science* não apresentou a informação completa para esses itens.

## 2.2.2 Etapa 2: Apresentação e interrelação dos dados

A pesquisa com os termos ("linked data") AND ("Open Educational Resources") retornou apenas 28 documentos. A pouca quantidade indica que há um potencial enorme de evolução da área. Ao realizar um paralelo entre a busca realizada somente pelo termo ("Open Educational Resources"), apresentada na Justificativa, subseção 1.3, pode-se perceber que a área de Recursos Educacionais Abertos está sendo bem pesquisada, entretanto não há tantos documentos que relatam usar técnicas para conexão dos dados. Pouco mais de 2% dos 1350 documentos encontrados em ("Open Educational Resources") apresentam essa relação.

A seguir serão apresentados os dados detalhados da pesquisa de acordo com a metodologia apresentada na subseção 2.1.2. Pode-se perceber que algumas porcentagens ultrapassam 100%. Isso porque a plataforma *ISI Web of Science* considera, por exemplo, mais uma uma classificação de área para o mesmo artigo. Da mesma maneira, alguns documentos são publicados por autores de diferentes países.

### 2.2.2.1 Análise das revistas mais relevantes

De acordo com (Mariano, Cruz e Gaitán 2011), o primeiro passo para investigar o estado da arte de um tema é conhecer quais são revistas científicas mais utilizadas, considerando a linha de pesquisa. A Tabela 2.2 apresenta as 10 revistas mais bem ranqueadas.

Tabela 2.2: Revistas mais relevantes.

POSIÇÃO	NOME DA REVISTA	TOTAL DE CITAÇÕES	FATOR DE IMPACTO
1	IEEE Communications Surveys and Tutorials	18,995	23.700
2	IEEE Transactions on Pattern Analysis And Machine Intelligence	52,888	17.861
3	Information Fusion	6,409	13.669
4	Journal of Statistical Software	25,372	13.642
5	IEEE Wireless Communications	8,140	11.391
6	IEEE Transactions On Evolutionary Computation	15,581	11.169
7	Medical Image Analysis	9,028	11.148
8	IEEE Transactions on Cybernetics	17,681	11.079
9	Journal of Industrial Information Integration	592	10.615
10	International Journal Of Intelligent Systems	5,796	10.312

Fonte: Autoria própria com base em dados do *ISI Web of Science*.



Para realizar essa busca, utilizou-se o recurso *Journal Citation Reports (JRC)* disponível no *ISI Web of Science*. As categorias escolhidas foram *Computer Science, Artificial Intelligence; Computer Science, Information Systems; Computer Science, Interdisciplinary Applications; Education and Educational Research; Education, Scientific Disciplines e Engineering, Industrial*. O ano de referência foi 2019.

### 2.2.2.2 Evolução do tema ano a ano

Em relação às publicações, os registros apresentaram seu pico em 2015 com 7 publicações e desde então vem apresentando queda ano a ano, com destaque para 2017, ano que houve apenas 1 publicação envolvendo Recursos Educacionais abertos e Dados Conectados.

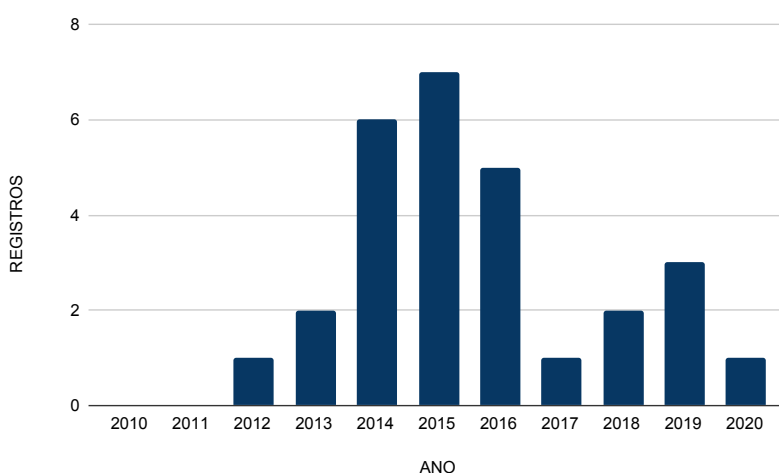


Figura 2.2: Publicações ano a ano sobre os termos Recursos Educacionais abertos e Dados Conectados.

Fonte: (Isotani e Bittencourt 2015).

### 2.2.2.3 Países que mais publicaram

A Tabela 2.3 apresenta os cinco países que mais publicaram. Pode-se notar que três, dos cinco países apontados como os que mais publicam, são europeus. O Equador aparece na segunda posição, sendo o país mais bem colocado da América Latina. O Brasil se apresentou na 9ª posição com 1 registro.

Tabela 2.3: Publicações por países.

PAISES	REGISTROS	% DE 28
SPAIN	14	50.000 %
ECUADOR	13	46.429 %
GERMANY	3	10.714 %
ITALY	3	10.714 %
AUSTRALIA	2	7.143 %

Fonte: Autoria própria com base em dados do *ISI Web of Science*.

#### 2.2.2.4 Organizações que mais publicaram

A fim de analisar a contribuição das universidades, registrou-se as cinco primeiras intuições para cada combinação. Os dados podem ser visualizados na Tabela 2.4. É interessante notar que 4 das 5 universidades pertencem aos países que mais publicam (Tabela 2.3). Esse fato comprova a relação direta entre de países que mais publicam e universidades que mais publicam.

Tabela 2.4: Organizações que mais publicaram

<b>ORGANIZAÇÕES</b>	<b>PAIS</b>	<b>REGISTROS</b>	<b>% DE 28</b>
UNIV TECN PARTICULAR LOJA	ECUADOR	10	35.714 %
UNIV POLITECN MADRID	SPAIN	9	32.143 %
LEIBNIZ UNIV HANNOVER	GERMANY	3	10.714 %
OPEN UNIV	ENGLAND	2	7.143 %
UNIV ALCALA DE HENARES	SPAIN	2	7.143 %

Fonte: Autoria própria com base em dados do *ISI Web of Science*.

#### 2.2.2.5 Áreas que mais publicam

As áreas que mais se destacaram foram: *Education Educational Research*, *Computer Science* e *Engineering*. Pode-se notar que essas áreas têm valores praticamente iguais de publicações, o que indica que os autores registram seus artigos com mais frequência em pelo menos duas dessas áreas, conforme mostra a Tabela 2.5.

Tabela 2.5: Áreas que mais publicam.

<b>ÁREA</b>	<b>REGISTROS</b>	<b>% DE 28</b>
EDUCATION EDUCATIONAL RESEARCH	14	50.000 %
COMPUTER SCIENCE	13	46.429 %
ENGINEERING	12	42.857 %
INFORMATION SCIENCE LIBRARY SCIENCE	2	7.143 %
TELECOMMUNICATIONS	2	7.143 %
ROBOTICS	1	3.571 %

Fonte: Autoria própria com base em dados do *ISI Web of Science*.

#### 2.2.2.6 Análise dos autores e artigos

A Tabela 2.6 apresenta os sete artigos mais citados. Esses artigos representam 68,18% de todas as citações. Os autores mais publicados foram: Nelson Piedra (12 registros), Janneth Chicaiza (11 registros), Jorge Lopez (7 registros), Edmundo Tovar (6 registros) e Edmundo Tovar-Caro (5 registros).

Pode-se analisar que Nelson Piedra aparece em 4 dos 7 artigos mais citados, Janneth Chicaiza e Jorge Lopez desenvolveram 3 artigos dos 7 mais citados. Por fim, Edmundo Tovar e Edmundo Tovar-Caro contribuíram com 2 cada. Esses dados reafirmam a relação direta entre autores que mais publicam e artigos mais citados. Há que se considerar também que a área é nova e esses trabalhos são pioneiros, datando de

Tabela 2.6: Artigos mais citados.

ARTIGO	AUTORES	ANO	CITAÇÕES	% DE 242
Interlinking educational resources and the web of data A survey of challenges and approaches	Dietze, Stefan Sanchez-Alonso, Salvador Ebner, Hannes Yu, Hong Qing Giordano, Daniela Marenzi, Ivana Nunes, Bernardo Pereira	2013	46	19,01%
Application of Resource Description Framework to Personalise Learning: Systematic Review and Methodology	Jevsikova, Tatjana Berniukevicius, Andrius Kurilovas, Eugenijus	2017	30	12,40%
Guest Editorial: Open Educational Resources in Engineering Education: Various Perspectives Opening the Education of Engineers	Tovar, Edmundo Piedra, Nelson	2014	22	9,09%
Seeking Open Educational Resources to Compose Massive Open Online Courses in Engineering Education An Approach Based on Linked Open Data	Piedra, Nelso; Chicaiza, Janneth Lopez, Jorge; Tovar, Edmundo	2015	17	7,02%
Open Data as Open Educational Resources: Towards transversal skills and global citizenship	Atenas, Javiera Havemann, Leo Priego, Ernesto	2015	16	6,61%
The Open University Linked Data - data.open.ac.uk	Daga, Enrico d'Aquin, Mathieu Adamou, Alessandro Brown, Stuart	2016	12	4,96%
Domain Categorization of Open Educational Resources Based on Linked Data	Chicaiza, Janneth Piedra, Nelson Lopez-Vargas, Jorge Tovar-Caro, Edmundo	2014	12	4,96%

### 2.2.2.7 Frequência de palavras chaves

De acordo com (Calazans, MASSON e MARIANO 2015), a análise das palavras-chave permite ao investigador entender melhor as linhas de pesquisa relacionadas a sua. Para isso, foi elaborada uma *Word Cloud* com as palavras-chaves extraídas dos 28 artigos encontrados. O diagrama foi feito por meio da ferramenta *TagCrowd*<sup>2</sup> utilizando as 50 palavras que tiveram mais frequência. Essa ferramenta permite uma melhor visualização, posto que as palavras com mais frequência aparecem com mais destaque.

As palavras que mais aparecem, sem considerar as palavras-chave buscadas dentro da pesquisa, relacionadas aos termos *Linked Data* e *Open Educational Resources*, foram: *Web*, *Learning*, *OER* e *Semantic*. Esse conjunto de palavras aponta o esforço em organizar os dados educacionais através da Web semântica

<sup>2</sup><<https://tagcrowd.com/>>

nos artigos pesquisados.

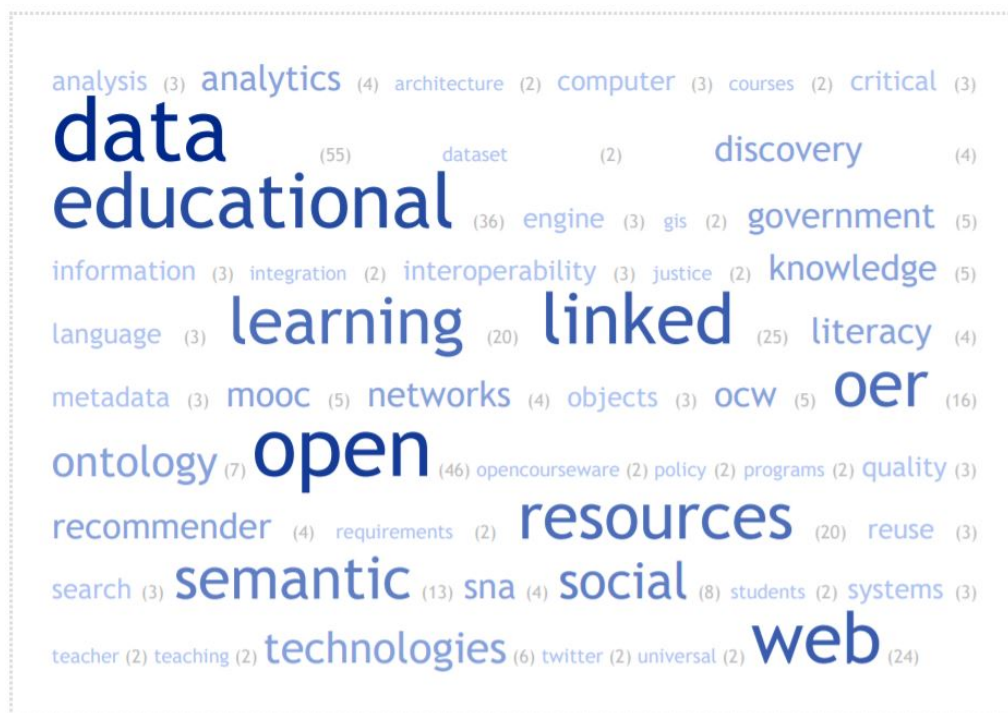


Figura 2.3: *Word Cloud* com as palavras-chaves dos artigos.

Fonte: (Isotani e Bittencourt 2015).

### 2.2.3 Etapa 3: Detalhamento, modelo integrador e validação por evidências

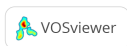
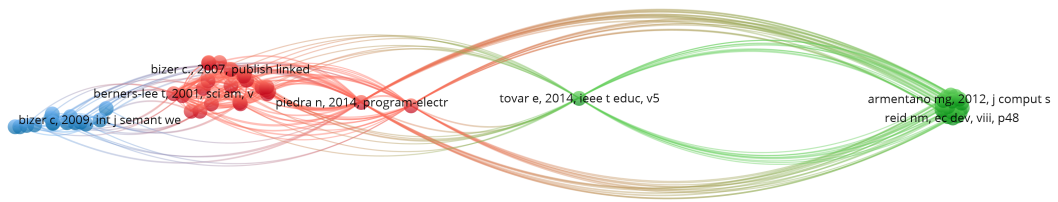
A última etapa do TEMAC é realizada para analisar as principais contribuições e abordagens relacionadas à pesquisa. Foi utilizado o *Software VOSviewer*<sup>3</sup> para elaborar mapas de calor a fim de visualizar melhor as mais relevantes ocorrências. Foram desenvolvidos três mapas: Co-citações, *Coupling* e Palavras mais frequentes.

#### 2.2.3.1 Co-citação

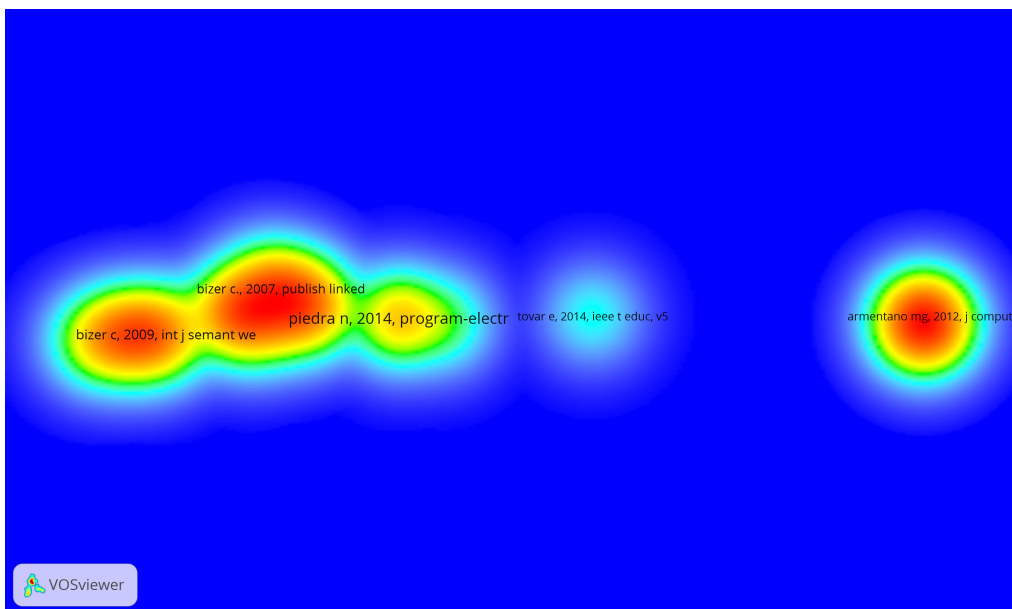
A análise da co-citação é o método bibliométrico mais comumente utilizado em análises bibliográficas e acontece quando dois ou mais artigos são citados juntos (Leung, Sun e Bai 2017). As Figuras 2.5(a) e 2.5(b) representam o Diagrama de Rede e os Mapa de Calor das co-citações encontradas no artigos, respectivamente.

A Figura 2.5(b) revelou três núcleos principais de co-citação: O núcleo mais forte diz respeito ao artigo de (Piedra, López e Chicaiza), que trata de um estudo de caso para a aplicação de Dados Abertos Conectados e está conectado a vários outros autores, como mostra a Figura 2.5(a). O segundo núcleo é dos artigos (Armentano, Godoy e Amandi 2012) e (Chang e Blei 2009) que tras o desenvolvimento de algoritmos para reduzir a análise de documentos por meio dos relacionamentos entre eles. Outro foco de

<sup>3</sup><<https://www.vosviewer.com/>>



(a) Diagrama de rede.



(b) Mapa de calor.

Figura 2.4: Co-citações encontradas no artigos.

Fonte: Autoria própria.

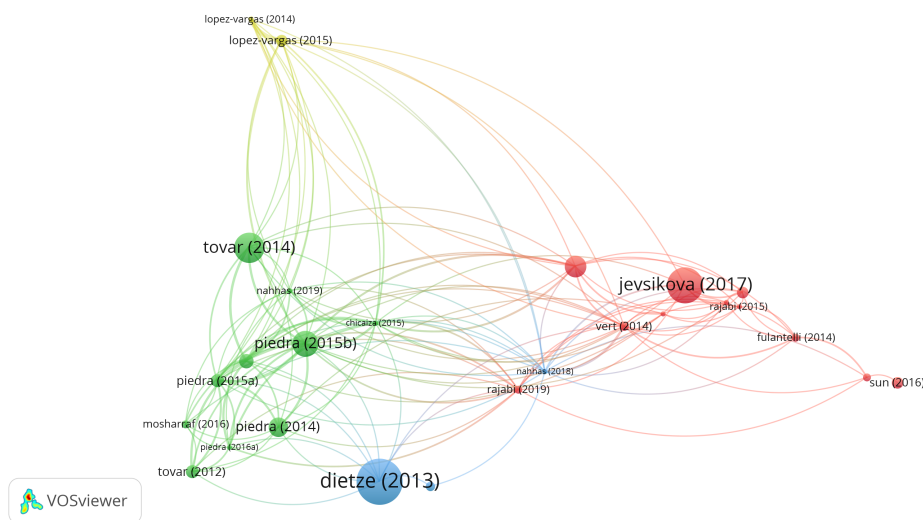
citações são os artigos de Bizer: (Bizer et al. 2007) e (Bizer, Heath e Berners-Lee 2011), que tratam sobre web Semântica e Dados Conectados.

Dos núcleos revelados, pode-se inferir que os artigos utilizam, em suas referências bibliográficas, arcabouços da Web Semântica e Inteligência Computacional.

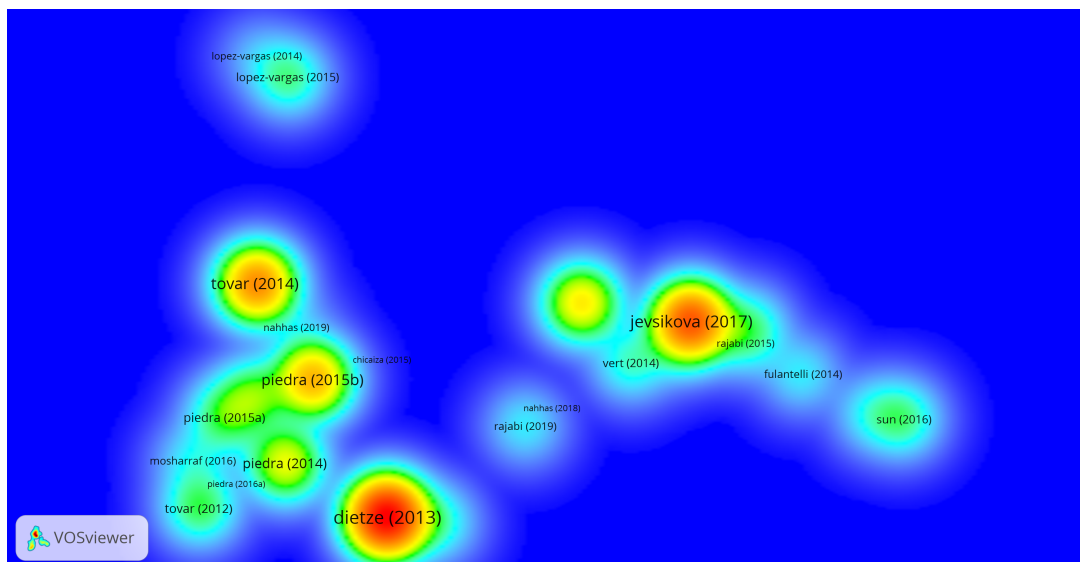
### 2.2.3.2 Acoplamento bibliográfico

(Vogel e Güttel 2013) afirmam que “acoplamento bibliográfico é uma medida de associação entre duas publicações citadas”. É estabelecido como uma leitura das referências feitas pelos autores e mostra as tendências de pesquisa mais fortes que estão sendo citadas.

Pode-se notar 3 núcleos principais: (Dietze et al. 2013), (Jevsikova et al. 2017) e (Tovar e Piedra 2014), conforme Figura 2.5. Os estudos realizados estão embasados em pesquisas bibliográficas sobre a aplicação de Dados Conectados em Aprendizagem aprimorada por tecnologia, desenvolvidos de um modelo para personalização de aprendizagem e estudos de caso sobre a aplicação de Recursos Educacionais Abertos no ensino da engenharia, respectivamente.



(a) Diagrama de rede.



(b) Mapa de calor.

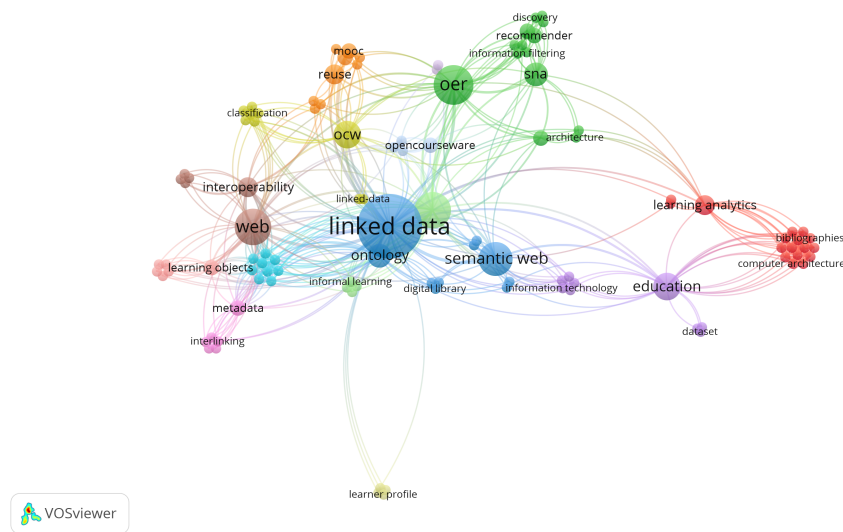
Figura 2.5: Acoplamento bibliográfico.

Fonte: Autoria própria.

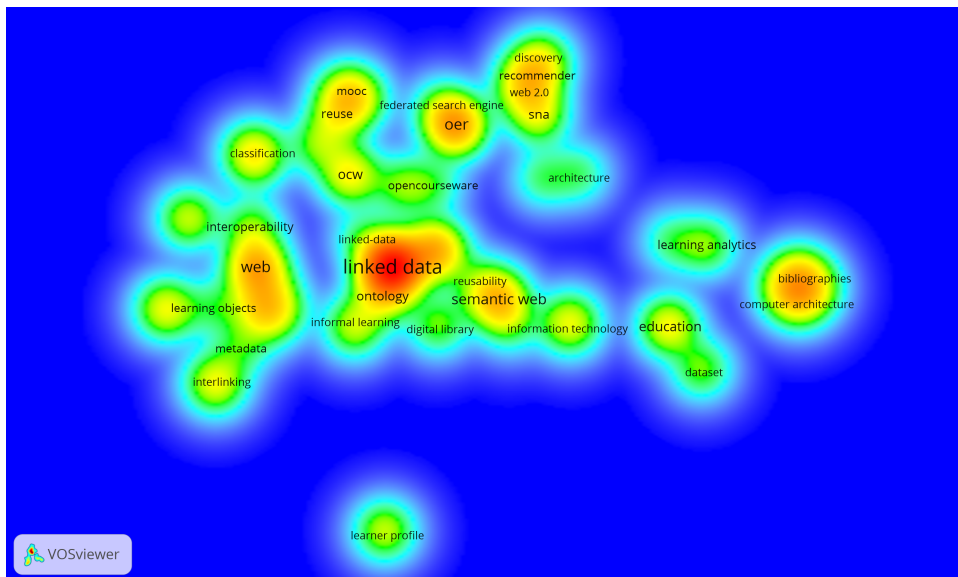
### 2.2.3.3 Palavras mais frequentes

A análise das palavras mais frequentes diz respeito à tendência do uso de palavras-chave no estudo, mas diferente do *Word Cloud*, mostrado na subseção 2.2.2.7, o Diagrama de Rede mostra a relação entre as palavras através de linhas que ligam os nós (Figura 2.6(a)). O Mapa de Calor traz as palavras mais citadas (tons de vermelho), além de sua relação com outras palavras, conforme Figura 2.6(a).

Pode-se notar que a palavra Ontologia e Web Semântica estão próximas do núcleo de Dados Conectados, Interoperabilidade e metadados estão próximos ao núcleo de Web e Conjunto de Dados (*dataset*) está próximo ao núcleo de educação.



(a) Diagrama de rede.



(b) Mapa de Calor.

Figura 2.6: Palavras mais frequentes.

Fonte: Autoria própria.

### 2.2.3.4 Determinação dos enfoques Teóricos de Pesquisa

Por fim, foi realizada uma avaliação dos enfoques dos artigos mais citados, conforme Tabela 2.7. Pode-se notar que a maioria dos artigos está pautada na aplicação dos conceitos da utilização de Dados Conectados em Recursos Educacionais Abertos.

Tabela 2.7: Enfoque dos artigos mais citados.

ARTIGO/AUTORES	ENFOQUE
Interlinking educational resources and the web of data A survey of challenges and approaches (Dietze et al. 2013)	Revisão bibliográfica sobre o uso de Dados Conectados relacionado à Aprendizagem aprimorada por tecnologia (TEL) e aplicação de padronizações para resolver problemas de interoperabilidade da área
Application of Resource Description Framework to Personalise Learning: Systematic Review and Methodology (Jevsikova et al. 2017)	Apresentação um <i>framework</i> para a personalização da aprendizagem baseada na aplicação do modelo padrão Resource Description Framework- RDF. A metodologia tem por base a aplicação de vocabulários considerados na área da aprendizagem
Guest Editorial: Open Educational Resources in Engineering Education: Various Perspectives Opening the Education of Engineers (Tovar e Piedra 2014)	O estudo apresenta seis experiências interessantes da aplicação de Recursos Educacionais Abertos ensino de engenharia
Seeking Open Educational Resources to Compose Massive Open Online Courses in Engineering Education An Approach Based on Linked Open Data (Piedra et al. 2015)	Revisão da literatura pautada em aplicação de tecnologias que apoiam a descoberta e inclusão de materiais educacionais abertos em grandes cursos online em educação de engenharia usando Dados Conectados e com foco no reutilização de dados.
Open Data as Open Educational Resources: Towards transversal skills and global citizenship (Atenas et al. 2015)	Apresentação de pesquisa sobre o uso de dados abertos no desenvolvimento de habilidades transversais
The Open University Linked Data - data.open.ac.uk (Daga et al. 2016)	Relata a evolução do data.open.ac.uk, a plataforma de Dados Abertos Conectados. A plataforma é um experimento de pesquisa e é um hub de dados para o conteúdo aberto da universitário
Domain Categorization of Open Educational Resources Based on Linked Data (Chicaiza et al. 2014)	Desenvolvimento de um processo de suporte à classificação semiautomática de Recursos Educacionais Abertos de acordo com as áreas temáticas utilizando Dados Conectados
Supporting openness of MOOCs contents through of an OER and OCW framework based on Linked Data technologies (Piedra et al. 2014)	Apresentação dos principais aspectos para a construção de uma estrutura baseada em tecnologias da web semântica para apoiar a inclusão de materiais abertos em cursos online massivos.



## 3 REFERENCIAL TEÓRICO

As instituições produtoras de conhecimento devem estar atentas para a divulgação dos recursos, posto que quanto mais os recursos são liberados, maiores as chances deles serem aproveitados pela sociedade. Entretanto se esses dados não estiverem aptos a interoperar com outros conjuntos de dados, o seu uso se torna bastante restrito. Dessa forma, qualidade dos dados gerados por pessoas e instituições produtoras de conhecimento está diretamente relacionada ao uso eficaz destes (Piedra et al. 2014).

A aplicação da Web Semântica em recursos educacionais busca transformar o uso desses dados, ao proporcionar maior facilidade na descoberta, utilização e compartilhamento dos dados, tornando-os interoperáveis com outros bancos de dados. Pode-se afirmar que a interoperabilidade de dados não é um "fim em si mesma". Ela permite incentivar benefícios sociais e tecnológicos para uma sociedade mais igualitária (Piedra et al. 2016).

Este Capítulo apresenta o embasamento teórico para o desenvolvimento do trabalho de pesquisa. A fundamentação teórica tem 4 pilares: Dados Conectados, Interoperabilidade de Dados, Recursos Educacionais Abertos e Qualidade de Dados.

### 3.1 DADOS CONECTADOS

A Web semântica foi criada com o objetivo de inovar a interação entre as informações disponíveis na internet, permitindo que não apenas humanos pudessem interpretar e conectar informações, mas também as máquinas, o que expande bastante as possibilidades de geração de valor dos dados. O conceito de Dados Conectados surgiu como uma proposta para padronização e publicação de dados na Web através de uso de critérios e protocolos internacionalmente recomendados. Recentemente o movimento de Dados abertos impulsionou a criação de Dados Conectados posto que os Dados Abertos tem como premissa que os dados sejam completos, primários, atualizados, acessíveis, processáveis por máquina e sem restrição de acesso por propriedade intelectual ou legislação. Dessa maneira, as padronizações dos Dados Conectados tem corroborado para a consolidação da Web semântica e o movimento Dados Abertos facilita a obtenção e o processamento de dados no ambiente virtual (Batista e Lóscio 2013).

A seguir serão apresentados conceitos importantes sobre Dados Abertos, Dados Conectados e Web Semântica.

### 3.1.1 Dados abertos

A produção de dados é inerente à sociedade, assim como a criação de novos dados baseados em já existentes (Isotani e Bittencourt 2015). Uma pesquisa divulgada pela (IDC Infobrief 2020) e patrocinada pela TIBCO aponta que 33 zettabytes de novos dados foram gerados em 2018. Desses dados gerados, apenas 14% são realmente novos, os outros 86% foram criados por replicação e distribuição. Os dados podem, assim, serem criados, coletados, armazenados, processados, distribuídos, consumidos e reciclados infinitamente (Floridi 2010), conforme ciclo apresentado na Figura 3.1. A pesquisa aponta, ainda, que há a projeção de criação de mais de 103 zettabytes de novos dados até 2023.



Figura 3.1: Ciclo de Vida da Informação.

Fonte: (Isotani e Bittencourt 2015).

Pode-se afirmar que a internet foi e continua sendo uma das causas fundamentais do crescimento de dados. Além dela, outro fator bastante relevante é a expansão do uso de dispositivos móveis. O uso da internet e dos dispositivos móveis não apenas corroboraram para a criação de dados, mas também para a facilitar o compartilhamento de informações (Marquesone 2016). De acordo com (Isotani e Bittencourt 2015), “este incrível crescimento de dados faz com que seja fundamental uma cuidadosa análise desses dados e a compreensão sobre quais tipos de dados têm sido gerados nesta sociedade da informação”.

Os dados são o maior bem da sociedade moderna. Entretanto, muitos destes não estão disponíveis para a geração de conhecimento e quando estão, se apresentam em formato que prejudica a sua utilização, posto que podem ser manipulados por homens e não, facilmente, por máquinas (Wood et al. 2014).

Com o propósito de impulsionar a democratização do acesso a dados, diversas instituições têm promovido o conceito de Dados Abertos. De acordo com a *Open Knowledge Foundation*, uma organização sem fins lucrativos que busca promover o conhecimento livre, “dados são abertos quando qualquer pessoa pode livremente usá-los, reutilizá-los e distribuí-los, estando sujeito a, no máximo, a exigência de creditar a sua

autoria e compartilhar pela mesma licença”<sup>1</sup>.

A primeira menção ao termo Dados Abertos está associada a um documento de uma agência científica americana em 1995. Este documento é pautado pela análise de dados sobre fenômenos naturais com base em comparações de dados de vários países. Em 2007, o termo criou forças quando entusiastas da internet se reuniram para conceitualizar Dados Abertos Públicos (Chignard 2013).

Os Dados Abertos Públicos, ou Governamentais, têm crescido entre os governos mundialmente. Segundo (Worthy 2015), “ao apoiar as reformas de dados abertos, os governos sinalizam seu compromisso de aumentar transparência, responsabilidade e participação”. (Coughlan 2020) ressalta que, embora o compartilhamento de dados na web não seja algo novo, o movimento de dados abertos gerou uma atmosfera que promove a amplitude e profundidade dos dados divulgados.

Segundo (World Bank 2019), “as primeiras políticas governamentais sobre Dados Abertos começaram a aparecer em 2009”. Um marco foi a criação da plataforma *Data.gov*<sup>2</sup> pelo governo Obama. Nesse portal, qualquer pessoa poderia ter acesso a dados abertos do governo americano (Isotani e Bittencourt 2015). Atualmente mais 250 governos, em nível nacional, subnacional e municipal, juntam esforços com entidades como Banco Mundial e as Nações Unidas para promover políticas públicas de dados abertos e há perspectiva de que cada vez mais países se juntem ao movimento (World Bank 2019).

Várias iniciativas foram criadas para incentivar os Dados Abertos Governamentais, entre elas a criação do *Open Government Partnership*, em 2011, quando líderes de governo, incluindo o Brasil, e defensores da sociedade civil se juntaram para criar uma organização que promovesse uma governança responsável, responsiva e inclusiva. Hoje com 78 membros, a organização trabalha com planos de ações bi-anuais. Cada membro elabora um plano de ação no qual estabelece metas para aumentar a transparência, responsabilidade e participação pública no governo (Open Government Partnership 2021).

Os líderes do G8 assinaram a Carta de Dados Abertos em 18 de junho de 2013. Esse documento, que continha cinco princípios básicos de como os governos deveriam utilizar seus dados para favorecer a transparência, inovação e responsabilidade, foi bem recebido por outros governos que apoiavam o governo aberto. Nos anos que se seguiram, a *Open Government Partnership*, aprimorou os princípios iniciais para deixá-los mais inclusivos e representativos, incluindo uma série de grupos de diversos países. A Carta Internacional de Dados Abertos (ODC) foi lançada na Assembleia Geral das Nações Unidas de 2015. Atualmente a Carta é adotada por 77 governos e endossada por 70 organizações (Open Data Charter 2020). Essa Carta, desenvolvida com o apoio de governos, sociedade civil e especialistas ao redor do mundo, estabelece normas para publicação de dados abertos (Open Data Charter 2015, FGV DAPP 2020). Essas normas são conhecidas como os 6 Princípios:

1. **“Open by default”**: Os dados podem ser acessados sem a necessidade de solicitá-los. Todos os dados devem estar disponíveis, exceto aqueles que comprometam o direito à privacidade, como dados de segurança da sociedade ou do estado;
2. **Oportunos e compreensíveis**: Dados são valiosos apenas se são relevantes. Publicar as informações de forma rápida e completa é essencial para explorar seu potencial. Ademais, os dados devem ser,

---

<sup>1</sup><<https://dados.gov.br/>>

<sup>2</sup><<https://www.data.gov/>>

na medida do possível, primários, ou seja, liberados em seu formato original;

3. **Acessível e utilizável:** Deve-se assegurar de que os dados possam ser lidos por máquinas e sejam fáceis de serem encontrados. Além disso, os dados devem possuir uma licença aberta e estarem em formatos de arquivos que facilitem a manipulação desses;
4. **Comparável e interoperável:** Para que os dados sejam mais eficientes e úteis, deve ser possível compará-los dentro de uma mesma organização e fora dela. Para isso, os dados devem estar em formato não estruturados e padronizados para apoiar a sua interoperabilidade, rastreabilidade e reutilização de forma eficaz;
5. **Para melhorar a governança e a participação social:** Os dados abertos tem o potencial de permitir que os cidadãos e outros que tenham interesse possam acompanhar melhor as atividades do governo, o que gera maior transparência e auxilia na melhoria de serviços públicos;
6. **Para desenvolvimento inclusivo e inovação:** Os dados abertos podem estimular a criatividade e promover a inovação, gerando empregos e renda. Além disso, o movimento de Dados Abertos proporciona uma maior visão sobre problemas sociais, sendo os dados utilizados para enfrentar desafios globais como a pobreza, as mudanças climáticas e a desigualdade, por exemplo.

Outra importante iniciativa foi a criação do *Open Data Barometer*<sup>3</sup>. Esta plataforma tem como propósito mensurar a continuidade e o verdadeiro impacto dos dados abertos no mundo por meio de análises de tendências globais. As análises são feitas através de indicadores que combinam dados contextuais, avaliações técnicas, entre outros indicadores secundários e possibilitam a comparação entre governos e regiões. Desde 2013, a *Open Data Barometer* lança relatórios anuais com análises de mais de 100 governos, entretanto em seu último relatório, elaborado em 2018, foi feita uma edição especial chamada de *Open Data Barometer - Leaders Edition*, que teve como foco os países que adotaram a Carta de Dados Abertos<sup>4</sup> ou que, sendo membro do G20, assinaram Princípios de Dados Abertos Anticorrupção<sup>5</sup>. No total, 30 países foram analisados (World Wide Web Foundation 2018). A análise tem como fatores macro de avaliação a prontidão, a implementação e os impactos dos dados abertos. Os estudos concluíram que houve uma considerável melhora dos países avaliadores em relação à primeira edição da análise, cinco anos antes. Não obstante, há uma disparidade entre o ritmo de evolução dos países e o desenvolvimento destes. No estudo, os países podem ser classificados em três grupos distintos:

- **Campeões:** são países que possuem maiores pontuações, todos acima de 65 pontos, o que indica que eles têm equilíbrio entre as categorias avaliadas e implementação eficaz dos dados abertos;
- **Competidores:** o segundo grupo, apesar de apresentar bons indicadores, ainda carecem de evidências de impactos nos critérios usados para a construção da pontuação, não conseguindo alcançar os 65 pontos para serem considerados do primeiro grupo;
- **Retardatários:** o último grupo mostra sérias deficiências em relação a pelo menos um dos critérios de avaliação e apresenta poucas melhorias em relação à primeira avaliação, em 2013.

---

<sup>3</sup><<https://opendatabarometer.org/>>

<sup>4</sup><<https://opendatacharter.net/>>

<sup>5</sup><<http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf>>

A Figura 3.2 apresenta o mapa dos 30 países em relação a classificação obtida, sendo que os governos que apresentam tons de verde são campeões, os que têm tons de amarelo correspondem aos países competidores e os que têm tons de vermelho são os retardatários. O Brasil obteve 50 pontos no estudo, sendo classificado como competidor.

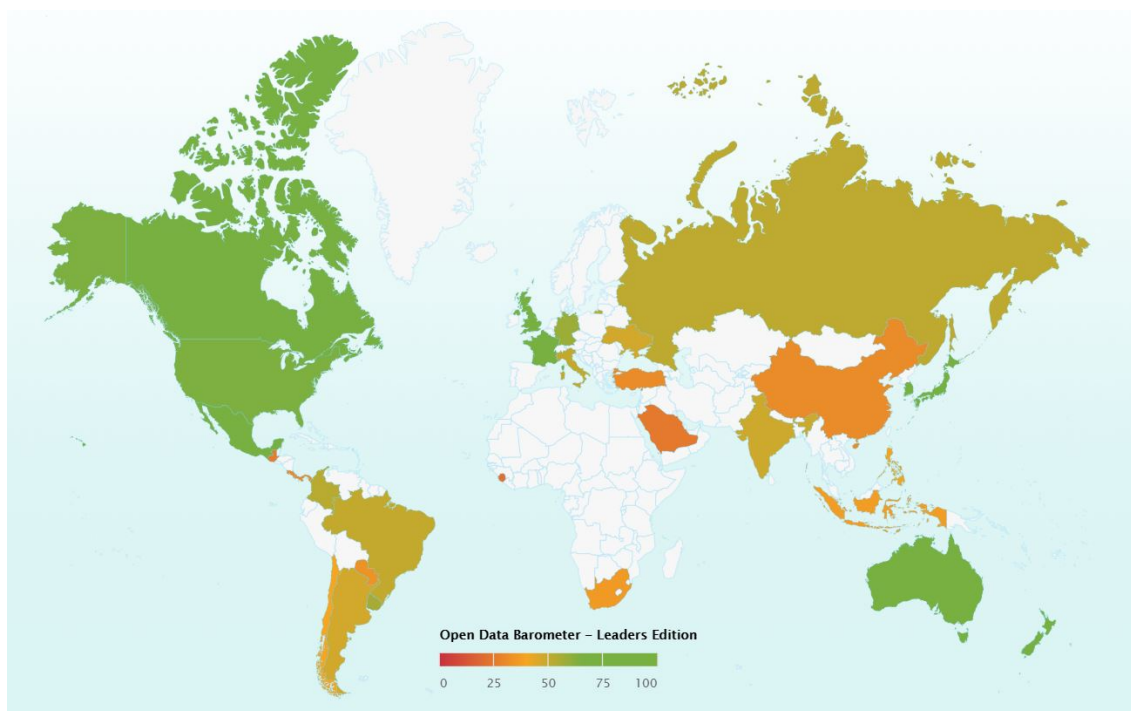


Figura 3.2: Classificação dos países avaliados no *Open Data Barometer - Leaders Edition*.

Fonte: (World Wide Web Foundation 2018)

A Tabela 3.1 mostra os principais acontecimentos, a nível mundial, vinculados à abertura de Dados Governamentais.

Tabela 3.1: Marcos da abertura de Dados Governamentais.

<b>ANO</b>	<b>ACONTECIMENTO</b>
2009	Lançamento das diretrizes de Governo Aberto pelo presidente Obama
2011	Criação do <i>Open Government Partnership</i>
2013	Criação do <i>Open Data Barometer</i>
2013	Assinatura do Carta de Dados Abertos pelo G8
2015	Lançamento da Carta Internacional de Dados Abertos na ONU

Fonte: Autoria própria.

Considerado um dos países pioneiros em relação a Dados Abertos Governamentais, o Brasil lançou, em 2014, o Portal da Transparência<sup>6</sup> com o objetivo de ser uma plataforma onde o cidadão possa obter informações sobre como o dinheiro público está sendo investido. Ademais, pode-se instruir sobre a gestão pública do país. A plataforma é, assim, um instrumento de controle social (Portal da Transparência 2018). No ano de 2011, além de ter fundado a Open Government Partnership, juntamente com outros países, o

<sup>6</sup><http://www.portaltransparencia.gov.br/>

Brasil aprovou a Lei de Acesso à Informação (LAI)<sup>7</sup>. Essa lei, em seu artigo 8º, torna obrigatório aos órgãos e entidades públicas a divulgação de dados, sem necessidade de requerimento, devendo estes estarem em local de fácil acesso, em formato aberto e não proprietário, de forma a possibilitar o acesso automático e legível por máquina. Em 2012, o país criou o Portal Brasileiro de Dados Abertos<sup>8</sup>, que tem como propósito a disponibilização de informações públicas pelo governo. O portal também objetiva ser um meio de contato entre a sociedade e o governo de forma a permitir o melhor aproveitamento dos dados para promover o desenvolvimento social e econômico do país. Nele, cada conjunto de dados (*dataset*, no inglês) está logicamente referenciado na Web e são publicados, por exemplo, dados do sistema de transporte, segurança pública e indicadores educacionais (Portal Brasileiro de Dados Abertos 2015).

Em 2016, através do Decreto nº 8.777/2016, foi instituída a Política de Dados Abertos do Poder Executivo Federal e se definiu a Infraestrutura Nacional de Dados Abertos (INDA) com propriedades normativas e gerencias (Governo Digital 2019). Desenvolvido pela Secretaria de Logística e Tecnologia da Informação – SLTI do Ministério do Planejamento, Orçamento e Gestão – MP, a INDA é “um conjunto de padrões, tecnologias, procedimentos e mecanismos de controle necessários para atender às condições de disseminação e compartilhamento de dados e informações públicas no modelo de Dados Abertos”, de acordo com as especificações do Padrões de Interoperabilidade de Governo Eletrônico - ePING (Portal Brasileiro de Dados Abertos 2020).

A Secretaria de Logística e Tecnologia da Informação – SLTI desenvolveu a *Cartilha Técnica para Publicação de Dados Abertos no Brasil* que apresenta recomendações para formatos de dados abertos (Portal Brasileiro de Dados Abertos 2019). Segundo o documento, “a filosofia de dados abertos não define exaustivamente uma lista de formatos permitidos. Para ser considerado um dado aberto, o conjunto de dados deve estar disponível em um formato de especificação aberta, não proprietário, e estruturado, ou seja, que possibilite seu uso irrestrito e automatizado através da Web”. Ademais, o documento recomenda que o formato seja amplamente conhecido. O SLTI alerta que um erro comum cometido pelas instituições é a publicação de dados em formato PDF, entretanto esse formato é não estruturado, o que dificulta e até mesmo inviabiliza a utilização dos dados. Recomenda-se, como formatos não proprietários e estruturados de dados abertos: JSON (*JavaScript Object Notation*), XML (*Extensible Markup Language*), CSV (*Comma-Separated Values*), ODS (*Open Document Spreadsheet*) e RDF (*Resource Description Framework*).

O movimento de Dados Abertos promete impulsionar cada vez mais a divulgação de informações, gerando valor para a sociedade. Entretanto, apesar dos esforços nacionais e internacionais de padronização, ainda muito dados são gerados de forma não padronizados, o que impossibilita a conexão entre estes. (Penteado 2020) afirma que “essa limitação restringe o potencial de reuso das informações, pois dificulta o cruzamento de informações de diferentes fontes, já que o processo de conhecimento das bases, limpeza dos dados, descoberta de conexões e interligação com outras bases de dados fica a cargo de quem os consome”. A fim de entender melhor as barreiras da abertura de dados, (Janssen, Charalabidis e Zuiderwijk 2012) realizaram entrevistas e um workshop e levantaram alguns impedimentos para a implementação dos dados abertos e dividiram em 6 categorias: Institucional, Complexidade da tarefa, Uso e participação, Legislação, Qualidade da Informação e barreiras técnicas.

---

<sup>7</sup><<http://www.planalto.gov.br/>>

<sup>8</sup><<https://dados.gov.br/>>

A abertura de dados beneficia diversas áreas de atuação como, por exemplo, a inovação científica. A pesquisadora (Marquesone 2016) ressalta que, com a liberação de dados “cientistas estão conseguindo acelerar o desenvolvimento de pesquisas tendo acesso a bases de dados que, até então, eram difíceis ou onerosas para se obter”. Outra área que tem se beneficiado com os dados abertos é o setor público. Na publicação *5 motivos para a abertura de dados na Administração Pública*, lançada em 2015, o Tribunal de Contas da União (TCU) registra 5 razões para a abertura de dados governamentais, sendo estas: transparência na gestão pública, contribuição da sociedade com serviços inovadores ao cidadão, aprimoramento na qualidade dos dados governamentais, viabilização de novos negócios e obrigatoriedade por lei (Portal Brasileiro de Dados Abertos 2015). O *Modelo de Referência de Abertura de Dados Abertos* desenvolvido pela (FGV DAPP 2020) cita, ainda, a economia de tempo para atendimento a pedidos por dados e informações governamentais, promoção de benefícios socioambientais e desenvolvimento sustentável e promoção do desenvolvimento econômico e social do país.

Para estimular a sociedade a utilizar dados abertos, Tim Berners-Lee, inventor da Web, criou um sistema de 5 estrelas em 2010. Este sistema apoia a identificação do nível de abertura que a organização se encontra em relação aos seus dados e contribui com passos estruturados para alcançar o nível mais alto, Dados abertos Conectados (Portal Brasileiro de Dados Abertos 2019).

Para cada estrela, estipulou-se formatos e níveis de conexão a serem ativados, conforme descrição a seguir (Berners-Lee 2010, Rautenberg et al. 2017, Penteadó 2020):

- **1ª Estrela:** A primeira estrela é atribuída quando os dados estão publicados na Web, não importando o formato. Para serem considerados dados abertos de primeira estrela, estes precisam apenas serem publicados sob uma licença aberta (Open License - OL). São exemplos comuns de dados de primeira estrela aqueles no formato de PDF ou imagem.
- **2ª Estrela:** Os dados com segunda estrela são aqueles que estão em um formato estruturado legível por máquina (Readable Machine - RE). Estes dados podem ser trabalhados por um software como as planilhas Excel, por exemplo, ou ainda serem convertidos em outros formatos.
- **3ª Estrela:** Dados com três estrelas possuem o formato aberto não proprietário (Open Format - OF) como, por exemplo, XML ou CSV. Devido a essa característica, sua manipulação pode ser feita através de um software com licença livre;
- **4ª Estrela:** A quarta estrela é concedida quando os dados usam Identificadores Uniformes de Recursos (Uniform Resource Identifier - URI). Esse recurso permite nomear os dados, facilitando a conexão entre eles.
- **5ª Estrela:** Na última fase dos dados abertos, estes estão conectados (Linked Data - LD) a fontes externas. Esse tipo de conexão permite uma contextualização mais ampla dos significados dos dados.

A Figura 3.3 representa o sistema, assim como níveis de esperados para os documentos.

É relevante ressaltar que, de acordo com a classificação de 5 estrelas, apenas se considera Dado Conectado se o dado tem pelo menos 4 estrelas, ou seja, se ele é identificado por meio de uma URI (Isotani e Bittencourt 2015).

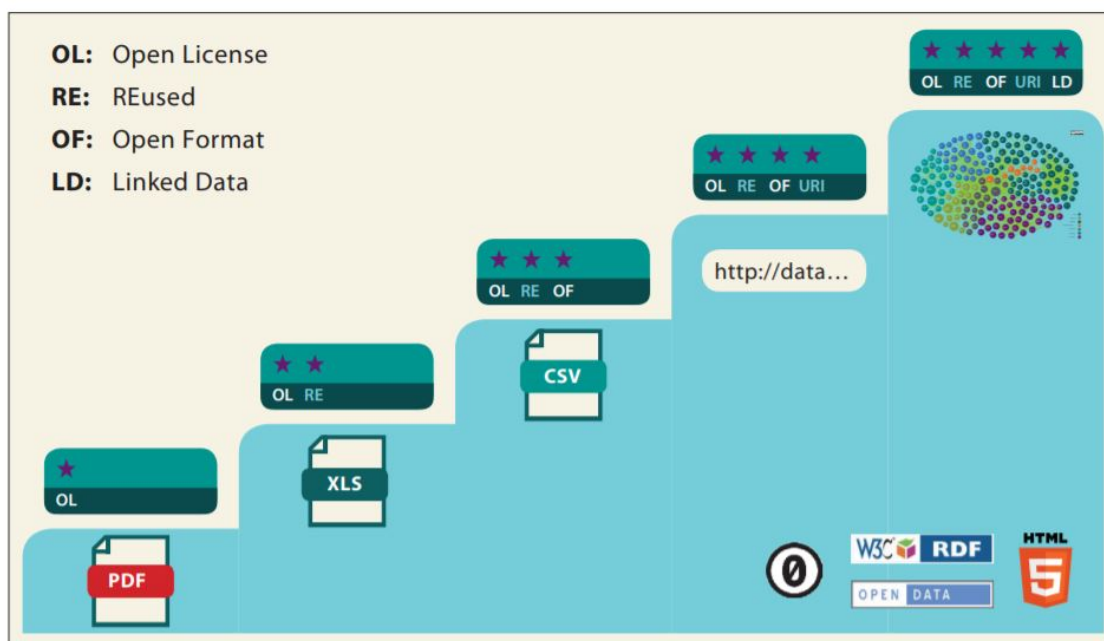


Figura 3.3: Sistema de 5 estrelas para dados abertos.

Fonte: (Isotani e Bittencourt 2015). Adaptado de (Berners-Lee 2010)

### 3.1.2 Web Semântica

A invenção da Web revolucionou o mundo. Criada em pela físico inglês Tim Berners-Lee em 1989 com objetivo de compartilhar informações no Centro Europeu de Pesquisas Nucleares, onde trabalhava, a inovação contou com três tecnologias: o HTML (*Hypertext Markup Language*), o servidor HTTP (*Hyper-text Transfer Protocol*) e o URI (*Unified Resource Identifier*). A primeira conferência internacional sobre World Wide Web (WWW) aconteceu em 1994 e marcou a ampliação da Web. Nesse evento, foi anunciada a criação do W3C, consórcio responsável por estabelecer padrões e desenvolver tecnologias para a Web (Isotani e Bittencourt 2015).

Na Web clássica, conhecida por Web de documentos, usa-se os hiperlinks para navegar entre as páginas, de modo que a linguagem HTML utiliza links e nós como pontos de conexão com outros documentos. Não obstante, nessa estrutura não permite a diferenciação entre os links, sendo assim, as máquinas não têm a capacidade de diferenciar significados entre os documentos Web. De acordo com (Bittencourt et al. 2008), isso causa um grande problema posto que, com a quantidade de dados produzidos pela humanidade, torna-se bastante difícil o uso eficaz de tal informação uma vez que pessoas não conseguem processar tantos dados em tempo hábil para que esses dados tenham valor. Essa problemática tem como solução a criação da Web Semântica ou Web de Dados. (Czerkawski 2014) afirma que “a Web Semântica apresenta uma solução para esta sobrecarga de informação, organizando as informações de maneira significativa para que possamos recuperar o que queremos sem nos perder na web”.

O termo Web Semântica foi usado pela primeira vez no livro *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web* escrito por Tim Berners-Lee em 1999. Porém, foi a publicação, em 2001, do artigo *The Semantic Web: a new form of Web content that is meaningful to computer*



*will unleash a revolution of new possibilities* (Berners-Lee, Hendler e Lassila 2001) que o movimento se fortaleceu e começou a ser pesquisado por cientistas globalmente. No artigo, os autores caracterizam e descrevem os princípios que possibilitam o funcionamento da Web semântica. (Isotani e Bittencourt 2015).

(Berners-Lee, Hendler e Lassila 2001) definiram a Web Semântica como “não uma Web separada, mas uma extensão da atual, em que a informação recebe um significado bem definido, permitindo melhor que computadores e pessoas para trabalharem em cooperação”. Dessa forma, a Web Semântica é uma evolução da Web clássica, possibilitando que máquinas possam navegar pela Web e encontrar significado entre as páginas e, assim, desenvolver tarefas complexas que antes somente poderiam ser feitas por pessoas. Para tal, a Web Semântica usa a Inteligência Artificial, a Engenharia de Software, a Computação Distribuída e outros recursos para que agentes de *software* possam interpretar informações (Isotani e Bittencourt 2015), fazendo que a utilidade dos dados aumentem ao se interligarem a outros dados (Penteado e Isotani 2020). A Web de dados faz uma contraposição a Web de documentos. Ao permitir as páginas web terem significado, a Web semântica possibilitou uma revolução no mundo virtual posto que a obtenção de valor nos dados passou a ser muito mais rápida e eficiente. (Penteado, Bittencourt e Isotani 2019) afirmam que a disponibilidade de dados em formato aberto que podem ser processados por máquinas facilita o cruzamento destes, favorecendo análises e aumentando a capacidade dos usuários na tomada de decisão em problemas difíceis.

Com o propósito de facilitar a implementação da Web Semântica, Tim Berners-Lee elaborou um modelo de camadas intitulado “Pirâmide da Web Semântica”, onde foram descritas as linguagens e recursos da Web de dados. Desde a criação do modelo original, a Web Semântica avançou rapidamente, sendo discutida em conferências, livros, periódicos e grupos de pesquisa, por exemplo. Devido a evolução e maturidade do conceito, em 2007 foram revisadas as tecnologias utilizadas de acordo com as recomendações do W3C (Isotani e Bittencourt 2015). A nova versão do Camadas da Web Semântica pode ser visualizada na Figura 3.4.

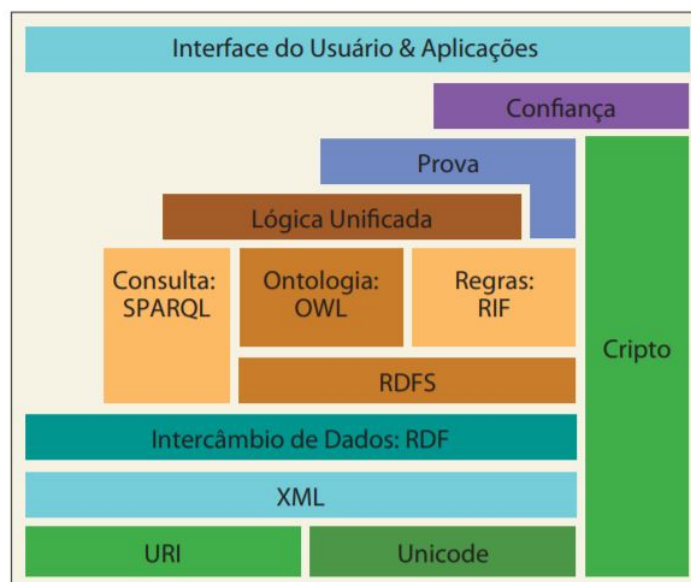


Figura 3.4: Camadas da Web Semântica Revisada.

Fonte: (Isotani e Bittencourt 2015). Adaptado de (Berners-Lee 2006)

Para que a Web de dados seja possível, é necessário algumas boas práticas como uso de vocabulários padronizados e de metadados. A seguir serão apresentados esses importantes conceitos.

### 3.1.2.1 Metadados

Metadados são definidos como dados sobre dados. Eles fornecem informações necessárias para o uso eficaz de um conjunto de dados, registrando informações como a hora, a data e o local de criação, por exemplo (World Bank 2019). Com uso de metadados estruturados é possível aumentar a interoperabilidade destes, uma vez que amplia os recursos de busca (FGV DAPP 2020).

Segundo (Laufer 2015), “os metadados têm como objetivo ajudar o processo de comunicação entre os publicadores e os consumidores de dados”. Dessa maneira, o uso de metadados estruturados é possível aumentar a interoperabilidade destes, uma vez que amplia os recursos de busca (FGV DAPP 2020).

### 3.1.2.2 Vocabulário e Ontologia

Os vocabulários definem “os conceitos e relacionamentos (também chamados de ‘termos’ ou ‘atributos’) usados para descrever e representar uma área de interesse” (W3C 2015). (Tomoyose 2021) afirma que “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”.

Os vocabulários têm um importante papel na definição de termos em conjuntos de dados, em especial para evitar ambiguidades de dados usados em diferentes áreas. De acordo com as recomendações do texto (Boas Práticas para Dados na Web (seção 4.1), usar vocabulários compartilhados e padronizados aumenta a interoperabilidade dos dados e corrobora com a comparação entre conjuntos de dados e o processamento automático destes (FGV DAPP 2020).

“Vocabulários padronizados devem ser reutilizados tanto quanto possível para facilitar a inclusão e expansão da rede de dados” (Dublin Core Metadata Initiative 2021). Assim, ao publicar dados, deve-se, sempre que possível, utilizar vocabulários existentes. Caso haja necessidade de criar um novo vocabulário é relevante utilizar o maior número de termos já reconhecidos a fim de evitar duplicação de conceitos (Laufer 2015).

Não há uma definição clara sobre o limite entre o que é considerado “vocabulário” e “ontologia”. Normalmente usa-se ontologia para conjunto de dados mais complexos e formalizados. Já o vocabulário é usado em termos mais vagos e que não necessitam de tanta formalização (W3C 2015). De acordo com (FGV DAPP 2020), escolher o nível de formação é importante para ajustar a necessidade da aplicação de forma a não gerar uma complexidade maior que a necessária.

Segundo (Fürber e Hepp 2010), as ontologias podem ser bastante benéficas ao se trabalhar com várias fontes de dados, apoiando na integração destes, por exemplo. (Isotani e Bittencourt 2015) ressalta que as ontologias “fornecem uma estrutura conceitual comum sobre a qual podemos desenvolver bases de conhecimento compartilháveis e reutilizáveis. E, em segundo lugar, facilitam a interoperabilidade e a fusão das informações”.

### 3.1.3 Estruturação de dados conectados

Os Dados Conectados estão no cerne do que a Web Semântica trata: integração em grande escala de dados na Web (Dietze et al. 2013, Piedra et al. 2015, W3C 2015). O conceito de Dados Conectados foi criado com para Tim Berners-Lee com o propósito de apoiar a padronização de dados dentro da Web. Desse modo, os dados Conectados “refere-se a um conjunto de práticas recomendadas para publicação de dados estruturados na web” (W3C 2016).

(Berners-Lee 2006) argumenta que a Web Semântica não é sobre colocar dados na Web. A Web de dados é sobre fazer links de forma que tanto pessoas como máquinas possam aproveitar os relacionamentos que existem. Com os Dados Conectados, ao pesquisar por um termo, pode-se encontrar todos aqueles relacionados com ele, assim, ao invés de conectar documentos HTML em um único espaço de informação, os Dados Conectados criam grafos de conexão global.

Enquanto a Web clássica se utiliza dos padrões HTML para acessar os documentos, a Web de dados utiliza os mesmos mecanismos de identificação e acesso que a Web clássica (URIs e HTTP), um modelo padrão de representação de dados, RDF (*Resource Description Framework*), e o SPARQL como uma linguagem de programação para acessar os dados. Podendo, assim, conectar e gerar relacionamentos entre os dados.

O RDF é um *framework* que apoia a representação de dados estruturados na web (W3C 2014). De acordo com (Curry et al. 2013), o RDF “fornece facilidades para mesclagem de dados, mesmo se os esquemas subjacentes forem diferentes. Também apoia a evolução de esquemas ao longo do tempo sem exigir que todos os consumidores de dados sejam alterados”.

#### 3.1.3.1 Formato do RDF

A base da estrutura da sintaxe do RDF é um uma relação entre um sujeito, um objeto e predicado conhecido como estrutura tripla. Ao conjunto dessas estruturas, dá-se-á o nome de Grafos RDF. Essa estrutura simples permite que dados estruturados e semiestruturados possam ser mesclados, expostos e compartilhados com diferentes bases de dados (Jevsikova et al. 2017). O grafo RDF pode ser entendido como um nó e um diagrama de arco direcionado, em que cada tripla representa uma conexão nó-arco-nó, conforme mostra a Figura 3.5 (W3C 2014).

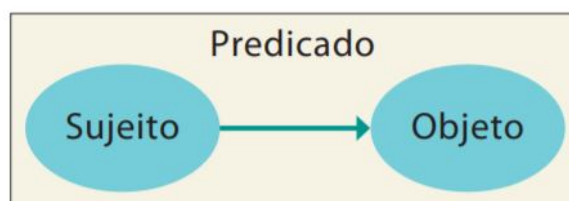


Figura 3.5: Gráfico RDF relacionando recursos.

Fonte: (Isotani e Bittencourt 2015). Adaptado de (W3C 2014).

No modelo RDF os recursos são quaisquer coisas que se deseja representar na web podendo ser concretos ou abstratos como, por exemplo, uma pessoa, um carro, um livro, um sentimento ou uma cor. O RDF representa a relação entre dois recursos (sujeito e objeto) e seu predicado, também chamado de propriedade (evidenciado pelo flecha) (Isotani e Bittencourt 2015, Adida et al. 2012). Essa representação sujeito, predicado e objeto é chamado de tripla

O exemplo da Figura 3.6 mostra o recurso livro se relacionando com outros três recursos: Gabriela, Cravo e Canela (propriedade título), 9788535912388 (propriedade isbn) e Jorge Amado (propriedade autor). Na Figura 3.7 é possível visualizar múltiplas triplas, comum em documentos RDF.

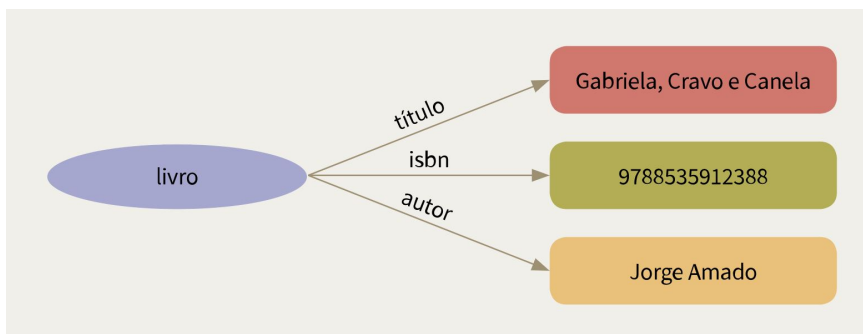


Figura 3.6: Exemplo de tripla RDF.

Fonte: (Laufer 2015)

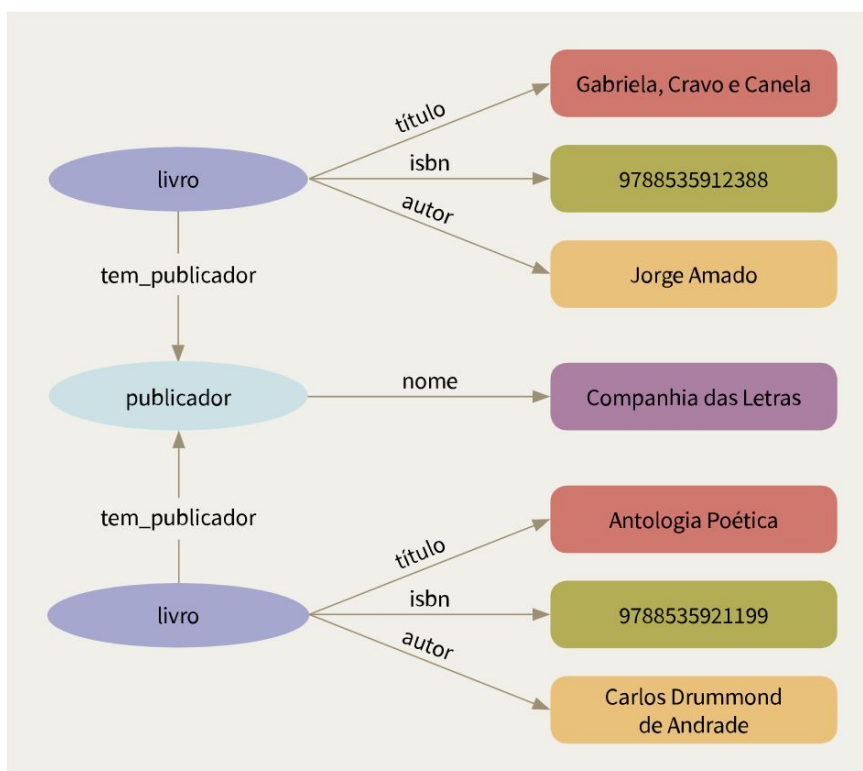


Figura 3.7: Múltiplas triplas RDF.

Fonte: (Laufer 2015)

Um fato importante é que o RDF foi criado para descrever recursos na Web. Para isso, o modelo usa de URIs, o que possibilita que cada recurso tenha uma identificação única e global na WWW (Laufer 2015, Isotani e Bittencourt 2015).

Para entender melhor o funcionamento dos URIs pode-se observar os Códigos 3.1 e 3.2. O Código 3.1 apresenta uma programação simples em HTML de um blog. Na página há o título e a data de criação da postagem. No Código 3.2 foram adicionados URIs que indicaram as propriedades título e data de criação por meio do vocabulário Dublin Core<sup>9</sup>. Através do URIs, as máquinas podem compreender o contexto da página, o que na Web dos documentos não é possível.

```
1 <html>
2 <head>
3   ...
4 </head>
5 <body>
6   ...
7   <h2>The Trouble with Bob</h2>
8   <p>Date: 2011-09-10</p>
9   ...
10 </body>
```

Código 3.1: Exemplo de código sem indicação de recursos . Fonte: (Adida et al. 2012).

```
1 <html>
2 <head>
3   ...
4 </head>
5 <body>
6   ...
7   <h2 property="http://purl.org/dc/terms/title">The Trouble with Bob</h2>
8   <p>Date: <span property="http://purl.org/dc/terms/created">2011-09-10</span></p>
9   ...
10 </body>
```

Código 3.2: Exemplo de código com URI identificados as propriedades título e data de criação. Fonte: (Adida et al. 2012).

### 3.1.4 Melhores Práticas para Publicação de Dados Conectados

Os autores (Hyland, Ateazing e Villazón-Terrazas 2014) elaboram uma recomendação W3C com o propósito de incentivar a publicação, acesso e utilização de Dados Abertos Governamentais, tendo como público alvo gestores, administradores e desenvolvedores web do sistema público. Neste contexto, dados abertos conectados permitem gerar maior transparência e responsabilidade social por parte dos governos (Open Knowledge Brasil 2017).

<sup>9</sup>O vocabulário Dublin Core será detalhado na seção 4.2.1.

As práticas estão listadas a seguir (Hyland, Ateazing e Villazón-Terrazas 2014, Isotani e Bittencourt 2015):

1. **Preparar os stakeholders:** A primeira etapa tem por objetivo preparar as partes interessadas em relação a criação e a manutenção do conjunto de Dados Abertos Conectados;
2. **Selecionar o conjunto de dados:** Na etapa seguinte, seleciona-se o conjuntos de dados que serão utilizados de acordo com os benefícios que podem gerar para as partes interessadas;
3. **Modelar os dados:** Na seguinte etapa, os dados são trabalhados de forma a representar recursos e serem utilizados em quaisquer aplicações;
4. **Especificar a licença:** Neste momento, ocorre a definição de qual será o grau de abertura dos dados e qual licença será utilizada;
5. **Nomear bons URIs:** Nessa prática deve-se estabelecer bons URIs, pensando na persistência dos identificadores. A publicação de dados por meio de URI é a base dos Dados Conectados;
6. **Usar vocabulários-padrões:** O uso de vocabulários padronizados facilita a conexão entre dados. Assim, na geração de dados, deve-se optar por vocabulários reconhecidos;
7. **Converter os dados:** Após a utilização de vocabulários, deve-se converter os dados em uma representação apropriada para Dados Conectados utilizando padrões RDF;
8. **Prover acesso aos dados:** Define-se quais as formas de acesso para humanos e agentes de *software*;
9. **Anunciar novo conjunto de Dados Conectados:** Etapa dedicada a divulgar o conjunto de dados para a sociedade com o propósito de impulsionar o uso desses dados;
10. **Reconhecer a função social:** Por fim, a última etapa está relacionada em reconhecer a responsabilidade de manter os dados na web ao longo do tempo.

Dados conectados não precisam necessariamente serem abertos. Há várias empresas que possuem dados conectados e fechados para uso de suas equipes, entretanto para serem considerados Dados Abertos Conectados, é necessário tenha licença aberta (Berners-Lee 2006).

## 3.2 INTEROPERABILIDADE DE DADOS

De acordo com (Geraci et al. 1991), na publicação *IEEE Standard Computer Dictionary*, interoperabilidade é a capacidade de dois sistemas de trocar informações, ou seja, a interoperabilidade pode ser entendida com a capacidade de um agente de software de se comunicar com outro, obtendo e mesclando dados.

A interoperabilidade de dados é obtida através da aplicação de padrões conceituados para possibilitar que as máquinas se comuniquem. Esses padrões podem ser objetos através do Conceito de Dados Conectados. (Curry et al. 2013) afirma que "Dados Conectados propõe uma abordagem para interoperabilidade de informação baseada na criação de um espaço global de informação."

(Ferreira e Segurado 2019) ressalta que a interoperabilidade “não está relacionada à integração clássica dos bancos de dados que tem como pressuposto um forte acoplamento dos dados”. Essencialmente, a interoperabilidade prevê um baixo acoplamento para compartilhamento dos dados e, portanto, a obtenção de uma metalinguagem comum para que esses dados sejam mutuamente inteligíveis nas análises.

### 3.3 RECURSOS EDUCACIONAIS ABERTOS

A primeira menção ao termo do Recursos Educacionais Abertos (do inglês *Open Educational Resources* - OER) aconteceu durante a convenção *Open Courseware for Higher Education in Developing Countries* realizada pela UNESCO em julho de 2002. Nessa convenção, a instituição reuniu um grupo de acadêmicos para discutir sobre um novo movimento: o OpenCourseWare (OCW), iniciado pelo *Massachusetts Institute of Technology* - MIT (Tovar e Piedra 2014, Piedra et al. 2015). Em 2007, a OCW Consortium foi criado e recentemente foi renomeado como *Open Education Consortium*<sup>10</sup>, sendo a maior organização internacional a promover os Recursos Educacionais Abertos (D’Antoni 2009).

A UNESCO define Recursos Educacionais Abertos (REA) como “materiais de ensino, aprendizagem e pesquisa em qualquer meio - digital ou outro - que residem no domínio público ou foram lançados sob uma licença aberta que permite acesso gratuito, uso, adaptação e redistribuição por outros com nenhuma ou restrições limitada”<sup>11</sup>. A Organização para a Cooperação e Desenvolvimento Econômico (OECD) define REA, ainda, como “materiais digitalizados oferecido livre e abertamente para educadores, alunos e autodidatas usarem e reutilizarem para ensino, aprendizagem e pesquisa” (OECD 2007)<sup>12</sup>.

Segundo (Chicaiza et al. 2014), “um dos principais objetivos do conhecimento aberto, e especificamente do movimento de Recursos Educacionais Abertos, é permitir que as pessoas acessem os recursos de que precisam para aprender”.

Uma pesquisa realizada pela McKinsey Global Institute analysis apontou que dados abertos aplicados à educação tem um potencial de gerar, aproximadamente, entre US\$890 bilhões a US\$1,2 trilhão por ano. Esse valor estaria, segundo a pesquisa, associado ao ensino mais eficaz dos alunos, à melhoria de instrução em sala de aula, o que ocasionaria a possibilidade de obtenção de melhores postos de trabalho e melhoria na tomada de decisão de pais e estudantes (McKinsey Global Institute analysis 2013).

Os dados abertos oferecem um acesso universal a registros digitais e análises para uma imensa gama de contextos. Isso pode criar avançadas possibilidades para atividades educacionais que ocorriam anteriormente apenas em ambientes acadêmicos ou onde se levaria muito tempo para serem realizadas sem o apoio das tecnologias web (Coughlan 2020).

Os pesquisadores (Bandeira et al. 2015) apresentam como resultado de suas pesquisas, quatro contextos que podem ser impulsionados pelo uso de dados abertos conectados aplicados à Educação: ensino, pesquisa, extensão e a gestão educacional, conforme apresentado a seguir:

---

<sup>10</sup><<https://www.oecconsortium.org/>>

<sup>11</sup><<https://en.unesco.org/themes/building-knowledge-societies/oer>>

<sup>12</sup><<http://www.oecd.org/education/ceeri/38654317.pdf>>

- Ensino:
  - Contribuição com o aprimoramento do material didático elaborado nas disciplinas, posto que os professores poderão reutilizar os conteúdos dos anos anteriores e atualizá-los no ano corrente. Ademais, pode-se enriquecer o conteúdo, apontando para recursos, como vídeos, áudios, entre outras tecnologias;
  - Apoio no processo de aprendizagem dos alunos, que poderão usufruir das matérias didáticas e registros de atividades de anos anteriores;
  - Fomento da interdisciplinaridade, que pode ser impulsionada por meio da divulgação de conteúdo de várias áreas. .
  
- Pesquisa:
  - Possibilidade de constituição e aprimoramento de pesquisas, uma vez que pesquisas relacionadas ficarão disponíveis de forma aberta em repositórios;
  - Facilidade na integração de pesquisas na mesma instituição e também de outras instituições, que terão acesso ao conteúdo mais facilmente;
  - Favorecimento da validação e procedência de referências bibliográficas, possibilitando a conexão entre as pesquisas.
  
- Extensão
  - Incentivo a novos projetos de extensão como continuação de projetos de pesquisas realizadas e disponibilizadas em repositórios;
  - Consolidação dos centros de inovação e incubadoras, posto que os dados podem ser usados em pesquisas e promoção de inovações tecnológicas nas empresas e outros centros de estudos;
  - Incentivo à inovação e geração de novos produtos e patentes por meio das pesquisas realizadas.
  
- Gestão
  - Contribuição com o acompanhamento das pesquisas realizadas dentro de empresas ou no meio acadêmico;
  - Possibilidade da sociedade acompanhar melhor as pesquisas realizadas dentro das universidades ao facilitar o acesso a esse tipo de informação;
  - Possibilidade de criar de novas ferramentas para melhoria da gestão;
  - Apoio na gestão do progresso de estudantes e das instituições de ensino.



### 3.4 QUALIDADE DE DADOS

De acordo com (Ciancarini, Poggi e Russo 2016), em trabalho apresentado no *2nd IEEE Second International Conference on Big Data Computing Service and Applications*, a garantia da qualidade de dados abertos é um dos maiores desafios daqueles que trabalham com a criação, utilização, reutilização e distribuição de dados. Uma pesquisa realizada pela (McKinsey Global Institute analysis 2013) aponta que 60% das organizações estão sendo desafiadas pela qualidade e complexidade de dados, sua integração, gestão e governança. Segundo (Zaveri et al. 2016) “a qualidade dos dados é crucial quando se trata de tomar decisões de longo alcance com base nos resultados de consultar vários conjuntos de dados”.

Os pesquisadores (Isotani e Bittencourt 2015) afirmam que para um dado aberto ser considerado de alta qualidade, ele precisa ser publicado, compartilhado e acessado por todos, permitindo que qualquer pessoa ou máquina possa encontrá-lo e manipulá-lo livremente. Além do mais, a formatação do dado deve possibilitar o cruzamento com outras fontes.

Dentro do ambiente acadêmico, diversas áreas têm abordado a qualidade de dados como, por exemplo, a engenharia, a gestão, a economia, a medicina, a geologia, a ciência da informação e a computação<sup>13</sup>.

Em seu estudo, (Wang e Strong 1996) realizam o que é considerada a primeira pesquisa empírica para a identificação de dimensões da qualidade. Para os autores, o conceito de qualidade de dados “é considerado multidimensional e as suas dimensões são definidas como um conjunto de atributos de qualidade que representam um único aspecto da qualidade de dados”(Fagundes, Macedo e Freund 2018).

De acordo com o modelo proposto por (Wang e Strong 1996), existem 4 grandes categorias que englobam 15 dimensões de qualidade de dados. As categorias são: Representacional, Contextual, Intrínsecas e Acessibilidade. (Zaveri et al. 2016) adaptou o modelo e adicionou mais quatro dimensões para incorporar as características dos Dados Conectados. As dimensões estão descritas a seguir:

- Acessibilidade
  - **Disponibilidade:** dados são considerados disponíveis se podem ser obtidos e estão prontos para uso;
  - **Licenciamento:** O licenciamento é uma nova dimensão que está relacionada aos Dados Conectados. Essa dimensão está vinculada ao fato dos dados necessitarem ter licença livre e de forma clara, tanto para seres humanos, quando para máquinas;
  - **Interligação:** Dimensão também relacionada com Dados Conectados e se refere às ligações entre conjuntos de dados, seja dentro de um mesmo conjunto ou externo a ele;
  - **Segurança:** A dimensão segurança se refere a quanto aos dados estão protegidos de alterações e uso indevido;
  - **Performance:** O desempenho se refere a eficiência do sistema que processa os dados e está relacionada ao tempo de resposta da busca dos dados.

- Intrínseca

---

<sup>13</sup><http://wcs.webofknowledge.com/>

- **Validade sintática:** A validade sintática é determinada pelo grau que o documento está de acordo com normas conceituadas;
  - **Precisão semântica:** A precisão semântica se relaciona a quanto os dados representam o mundo real;
  - **Consistência:** A consistência é a dimensão que trata da lógica dos dados. Para um conjunto de dados ser consistente, eles não podem ter contradições no que diz respeito à representação da informação;
  - **Integridade:** Refere-se a quanto às informações necessárias estão presentes dentro do conjunto de dados ou ao grau que um conjunto de dados está interligado;
  - **Concisão:** A concisão se refere a minimização das redundâncias dentro de um conjunto de dados, de forma que esse conjunto seja o mais reduzido possível e agregue o mesmo valor.
- Representacional
    - **Concisão representacional:** A Concisão representacional diz respeito a quanto os dados estão bem compactados e formatados. Um exemplo seria a geração de URL curtas para facilitar o processamento em larga escala;
    - **Interoperabilidade:** Interoperabilidade é o grau que o conjunto de dados está em conformidade com padrões estabelecidos pela comunidade relacionada a web para poder se conectar com outras fontes de dados;
    - **Interpretabilidade:** A interpretabilidade faz referência a quanto os dados seguem notações técnicas, possibilitando que as máquinas os processem;
    - **Versatilidade:** Um conjunto de dados versátil é aquele que pode ser disponibilizado em diferentes representações e formas padronizadas;
  - Contextual
    - **Relevância:** Relevância se refere aos dados estarem de acordo com as necessidade dos usuários, tendo importância em relação ao tema procurado;
    - **Confiabilidade:** Confiabilidade é tido como o grau em que as informações são aceitas como corretas e confiáveis;
    - **Compreensibilidade:** A compreensibilidade dos dados está relacionada com quão fácil eles podem ser entendidos sem ambiguidade;
    - **Pontualidade:** Essa dimensão se relaciona a quanto os dados estão atualizados em relação às necessidades dos consumidores;

A Figura 3.8 representa as 4 categorias e as 18 dimensões propostas por (Zaveri et al. 2016).

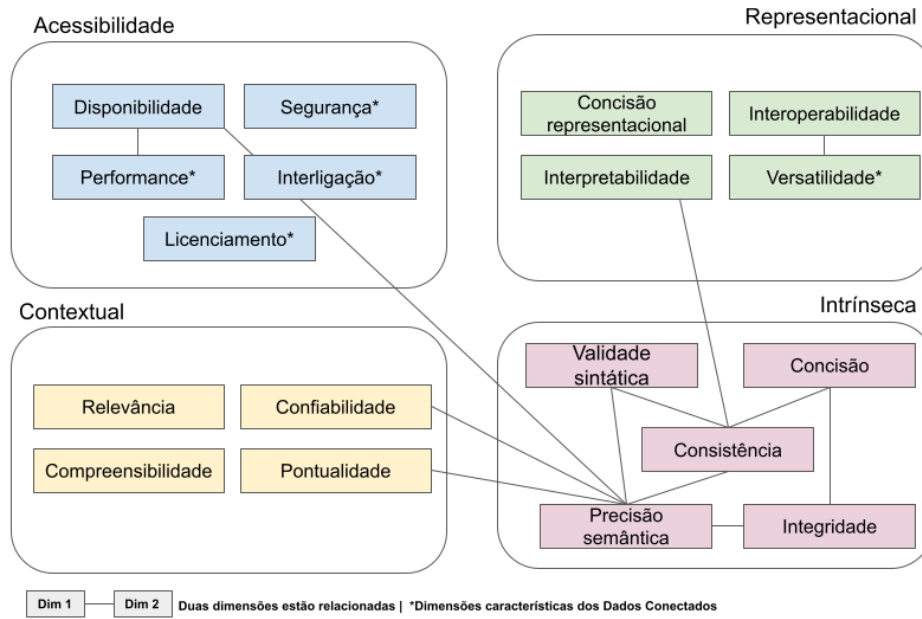


Figura 3.8: Dimensões da qualidade de dados.

Fonte: Adaptado de (Zaveri et al. 2016).

## 4 BOAS PRÁTICAS E PADRÕES DE INTEROPERABILIDADE E QUALIDADE DE DADOS

A expansão e flexibilização da Web trouxeram alguns desafios para publicadores e consumidores de dados como, por exemplo, dificuldades de acesso, compreensão e uso destes. Diferente de sistemas fechados, a Web pode apresentar diversas formas de configurações e disponibilização de conteúdos. Assim sendo, é essencial o uso de diretrizes que auxiliem aos publicadores a gerarem dados mais consistentes (Lóscio, Burle e Calegari 2017).

Este Capítulo apresenta boas práticas e padrões de representações sintáticas e de vocabulários que corroboram para a interoperabilidade e qualidade de dados na Web.

### 4.1 BOAS PRÁTICAS PARA DADOS NA WEB

os pesquisadores (Lóscio, Burle e Calegari 2017) desenvolveram um conjunto de boas práticas denominadas Boas Práticas para Dados na Web (do inglês, *Data on the Web Best Practices* - DWBP), sendo esta uma recomendação W3C. O documento objetiva a expansão da Web e promoção de maiores trocas de dados e o relacionamento entre eles.

As Boas Práticas estão relacionadas com conjuntos de dados e suas distribuições, sendo que uma distribuição representa “um determinado formato disponível de um conjunto de dados” como CSS ou API. As distribuições facilitam o compartilhamento de dados em larga escala, o que contribui para que os dados sejam utilizados por diferentes consumidores, sem importar a finalidade. O fornecimento de alguns dados e informações como metadados, vocabulários padronizados, procedências, versão e licença aumenta a confiabilidade e reuso dos dados. Outro importante fator é o uso da base arquitetônica da Web como, por exemplo, o uso de URIs para identificar recursos (Lóscio, Burle e Calegari 2017). A Figura 4.1 representa os principais elementos dos DWBP, assim como estes se relacionam.



Figura 4.1: Representação das interações entre os elementos de dados no DWBP.

Fonte: (Lóscio, Burle e Calegari 2017).

As Boas Práticas para Dados na Web foram desenvolvidos a partir de desafios listados no documento Requisitos e Casos de Uso das Boas Práticas para Dados na Web (do inglês *Data on the Web Best Practices Use Cases & Requirements*<sup>1</sup>). As Boas Práticas respondem a 12 questionamentos, possuindo 12 dimensões e 35 boas práticas associadas a elas (Lóscio, Burle e Calegari 2017, Penteado, Bittencourt e Isotani 2019). As dimensões e as boas práticas são apresentadas a seguir:

1. **Metadados** (como fornecer metadados para humanos e computadores?): fornecer metadados (BP1); fornecer metadados descritivos (BP2); fornecer metadados estruturais (BP3);
2. **Licença de Dados** (como permitir/restringir o acesso?): fornecer informações sobre as licenças para os dados (BP4);
3. **Proveniência e Qualidade** (Como aumentar a confiança?): fornecer informações sobre a proveniência dos dados (BP5); fornecer informação sobre a qualidade dos dados (BP6);
4. **Versionamento de Dados** (como encontrar diferentes versões e seus históricos?): fornecer indicador de versões (BP7); fornecer histórico de versões (BP8);
5. **Identificadores de Dados** (como identificar conjunto de dados e distribuições?): usar URIs persistentes como identificadores de conjunto de dados (BP9) e dentro de conjunto de dados (BP10); atribuir URIs para versões de *datasets* e séries (BP11);
6. **Formatos de Dados** (quais formatos de dados posso usar?): usar formatos padronizados, processáveis por computador (BP12); usar representações neutras de localidade (BP13); fornecer os dados em múltiplos formatos (BP14);
7. **Vocabulários de Dados** (como melhorar a interoperabilidade dos dados?): reusar vocabulários, de preferência os padronizados (BP15); escolher o nível adequado de formalização (BP16);
8. **Acesso a Dados** (como oferecer acesso aos dados?): fornecer download em massa (BP17); fornecer subconjuntos para grandes conjuntos de dados (BP18); usar negociação de conteúdo para oferecer dados em formatos múltiplos (BP19); fornecer acesso em tempo real (BP20); fornecer dados atualizados (BP21); fornecer explicação para dados que não estejam disponíveis (BP22); Disponibilizar dados por meio de API (BP23); usar padrões da Web como fundamento para a API (BP24); fornecer documentação completa para a API (BP25); evitar mudanças que possam afetar o API (BP26);
9. **Preservação de Dados** (como arquivar dados?): preservar os identificadores (BP27); avaliar a cobertura do conjunto de dados (BP28);
10. **Feedback** (como motivar os usuários?): coletar *feedback* dos consumidores dos dados (BP29); tornar o *feedback* disponível (BP30);
11. **Enriquecimento de Dados** (como acrescer valor aos dados?): enriquecer dados por meio da geração de novos dados (BP31); fornecer representações complementares dos dados (BP32);

---

<sup>1</sup><https://www.w3.org/TR/dwbp-ucr/>

12. **Replicação de Dados** (como reusar com responsabilidade os dados?): fornecer *feedback* para o publicador original (BP33); seguir os termos das licenças (BP34); citar a publicação original (BP35);

As recomendações DWBP apontam alguns benefícios trazidos pelos usos das Boas Práticas:

- **Compreensão:** os usuários poderão entender melhor a estrutura e o que significa os dados, metadados e procedência dos conjuntos de dados;
- **Facilidade de Processamento:** os agentes de *software* serão capazes de processar e utilizar automaticamente os *datasets*;
- **Facilidade de Descoberta:** Os agente de *software* poderão descobrir automaticamente conjunto de dados ou subconjuntos de dados;
- **Reúso:** serão amplificadas as possibilidades de reutilização de um *dataset* por variados grupos de usuários de dados;
- **Confiança:** aumentará a confiança deposita em um conjunto de dados;
- **Capacidade de Conexão:** possibilitará a criação de novas conexões entre itens ou conjuntos de dados;
- **Facilidade de Acesso:** os dados poderão ser mais facilmente acessados por pessoas e máquinas vários formatos diferente;
- **Interoperabilidade:** auxiliar no entendimento entre publicadores e consumidores de dados.

A Tabela 4.1 mostra os benefícios associados às Boas Práticas para Dados.

Tabela 4.1: Benefícios das Boas Práticas.

DIMENSÃO	BOA PRÁTICA		BENEFÍCIO
	Nº	Descrição	
Metadados	1	Fornecer metadados	Reuso, compreensão, descoberta e processabilidade
	2	Fornecer metadados descritivos	Reuso, compreensão e descoberta
	3	Fornecer metadados estruturais	Reuso, compreensão e processabilidade.
Licenças de dados	4	Fornecer informações sobre as licenças para os dados	Reuso e confiabilidade
Proveniência de dados	5	Fornecer informações sobre a proveniência dos dados	Reuso, compreensão e confiabilidade

**Tabela 4.1 continuação da página anterior.**

DIMENSÃO	BOA PRÁTICA		BENEFÍCIO
	Nº	Descrição	
Qualidade de dados	6	Fornecer informação sobre a qualidade dos dados	Reuso e confiabilidade
Versionamento de dados	7	Fornecer um indicador de versão	Reuso e confiabilidade
	8	Fornecer histórico de versões	Reuso e confiabilidade.
Identificadores de dados	9	Usar URIs persistentes como identificadores de conjuntos de dados	Reuso, ligação, descoberta e interoperabilidade.
	10	Usar URIs persistentes como identificadores dentro dos conjuntos de dados	Reuso, ligação, descoberta e interoperabilidade
	11	Atribuir URIs para versões de <i>datasets</i> e séries	Reuso, descoberta e confiabilidade
Formatos de dados	12	Usar formatos padronizados, processáveis por computador	Reuso e processabilidade
	13	Usar representações neutras de localidade	Reuso e compreensão
	14	Fornecer os dados em múltiplos formatos	Reuso e processabilidade
Vocabulários de dados	15	Reusar vocabulários, de preferência os padronizados	Reuso, processabilidade, compreensão, confiabilidade e interoperabilidade
	16	Escolher o nível adequado de formalização	Reuso, compreensão e interoperabilidade
Acesso aos dados	17	Fornecer download em massa	Reuso e acesso
	18	Fornecer subconjuntos para grandes conjuntos de dados	Reuso, ligação, acesso e processabilidade
	19	Usar negociação de conteúdo	Reuso e acesso
	20	Fornecer acesso em tempo real	Reuso e acesso
	21	Fornecer dados atualizados	Reuso e acesso
	22	Fornecer explicação para dados que não estejam disponíveis	Reuso e confiabilidade
	23	Disponibilizar dados por meio de API	Reuso, processabilidade, interoperabilidade e acesso
	24	Usar padrões da Web como fundamento para a API	Reuso, ligação, descoberta, acesso, processabilidade e interoperabilidade
	25	Fornecer documentação completa para a API	Reuso e confiabilidade

**Tabela 4.1 continuação da página anterior.**

DIMENSÃO	BOA PRÁTICA		BENEFÍCIO
	Nº	Descrição	
Preservação de dados	26	Evitar mudanças que possam afetar o API	Confiabilidade e interoperabilidade
	27	Preservar os identificadores	Reuso e confiabilidade
	28	Avaliar a cobertura do conjunto de dados	Reuso e confiabilidade
<i>Feedback</i>	29	Coletar <i>feedback</i> dos consumidores dos dados	Reuso, compreensão e confiabilidade
	30	Tornar o <i>feedback</i> disponível	Reuso e confiabilidade
Enriquecimento de dados	31	Enriquecer dados por meio da geração de novos dados	Reuso, compreensão, confiabilidade e processabilidade
	32	Fornecer representações complementares dos dados	Reuso, compreensão, acesso e confiabilidade
Republicação	33	Fornecer <i>feedback</i> para o publicador original	Reuso, interoperabilidade e confiabilidade
	34	Seguir os termos das licenças	Reuso e confiabilidade
	35	Citar a publicação original	Reuso, descoberta e confiabilidade

Fonte: Adaptado de (Tomoyose 2021).

Analisando o modelo de (Wang e Strong 1996), adaptado por (Zaveri et al. 2016), apresentado na Seção 3.4, pode-se fazer um paralelo entre as dimensões relacionadas à qualidade de dados e os algumas Boas práticas para Dados na Web, conforme mostrado na Tabela 4.2. As relações descritas na tabela foram desenvolvidas a partir das descrições das dimensões do modelo da qualidade de dados e dos benefícios advindos das Boas Práticas.

Tabela 4.2: Correlação entre dimensões das qualidade de dados Boas práticas para Dados.

CATEGORIA	DIMENSÃO	CONCEITO	BOA PRÁTICA	
			Nº	Descrição
Disponibilidade		Dados disponíveis e prontos para uso	20	Fornecer acesso em tempo real
			21	Fornecer dados atualizados
Licenciamento		Fornecimento de licença livre e de forma clara	4	fornecer informações sobre as licenças para os dados

Acessibilidade



Tabela 4.2 continuação da página anterior.

CATEGORIA	DIMENSÃO	CONCEITO	BOA PRÁTICA	
			Nº	Descrição
	Interligação	Conexão entre <i>datasets</i>	9	Usar URIs persistentes como identificadores de conjuntos de dados
			10	Usar URIs persistentes como identificadores dentro dos conjuntos de dados
			18	Fornecer subconjuntos para grandes conjuntos de dados
			24	Usar padrões da Web como fundamento para a API
	Segurança	Proteção contra alterações e usos indevidos	4	Fornecer informações sobre a licença de dados
	Performance	Tempo de processamento	1	Fornecer metadados
			3	Fornecer metadados estruturais
			12	Usar formatos padronizados, processáveis por computador
			14	Fornecer os dados em múltiplos formatos
			15	Reusar vocabulários, de preferência os padronizados
			18	Fornecer subconjuntos para grandes conjuntos de dados
			23	Disponibilizar dados por meio de API
	31	Enriquecer dados por meio da geração de novos dados		
Validade sintática	Conformidade com normas conceituadas	4	Fornecer informações sobre a licença de dados	
		15	Reusar vocabulários, de preferência os padronizados	
		34	Seguir os termos das licenças	
Precisão semântica	Representação fiel ao mundo real	1	Fornecer metadados	
		2	Fornecer metadados descritivos	
		3	Fornecer metadado estruturais	

Tabela 4.2 continuação da página anterior.

CATEGORIA	DIMENSÃO	CONCEITO	BOA PRÁTICA	
			Nº	Descrição
			5	Fornecer informações sobre a proveniência dos dados
			13	Usar representações neutras de localidade
			15	Reusar vocabulários, de preferência os padronizados
			16	Escolher o nível adequado de formalização
			29	Coletar <i>feedback</i> dos consumidores dos dados
			32	Fornecer representações complementares dos dados
	Consistência	Eliminação das contradições dentro do conjunto de dados	1	Fornecer metadados
			13	Usar representações neutras de localidade
			15	Reusar vocabulários, de preferência os padronizados
			16	Escolher o nível adequado de formalização
	Integridade	Informações necessárias estão dentro de <i>datasets</i> ou grau de interligações de <i>datasets</i>	1	Fornecer metadados
			2	Fornecer metadados descritivos
			3	Fornecer metadados estruturais
			15	Reusar vocabulários, de preferência os padronizados
			16	Escolher o nível adequado de formalização
			32	Fornecer representações complementares dos dados
	Concisão	Conjunto de dados seja o menor possível, agregando o valor esperado	16	Escolher o nível adequado de formalização
			29	Coletar <i>feedback</i> dos consumidores dos dados
33			Fornecer <i>feedback</i> para o publicador original	
Concisão representacional	Representação bem formatada e compactada dos dados	9	Usar URIs persistentes como identificadores de conjuntos de dados	

Tabela 4.2 continuação da página anterior.

CATEGORIA	DIMENSÃO	CONCEITO	BOA PRÁTICA	
			Nº	Descrição
			10	Usar URIs persistentes como identificadores dentro dos conjuntos de dados
	Interoperabilidade	Conformidade com padrões para poder se conectar a outras fontes de dados	9	Usar URIs persistentes como identificadores de conjuntos de dados
			10	Usar URIs persistentes como identificadores dentro dos conjuntos de dados
			15	Reusar vocabulários, de preferência os padronizados
			16	Escolher o nível adequado de formalização
			23	Disponibilizar dados por meio de API
			24	Usar padrões da Web como fundamento para a API
			26	Evitar mudanças que possam afetar o API
			33	Fornecer <i>feedback</i> para o publicador original
	Interpretabilidade	Conformidade com formatos padronizados, processáveis por computador	12	Usar formatos padronizados, processáveis por computador
			15	Reusar vocabulários, de preferência os padronizados
			19	Usar negociação de conteúdo para oferecer dados em formatos múltiplos
	Versatilidade	Dados disponíveis em diferentes representações e formas padronizadas	12	Usar formatos padronizados, processáveis por computador
			19	Usar negociação de conteúdo para oferecer dados em formatos múltiplos
	Relevância	Dados de acordo com a necessidade do usuário	16	Escolher o nível adequado de formalização
			29	Coletar <i>feedback</i> dos consumidores dos dados
			33	Fornecer <i>feedback</i> para o publicador original

Tabela 4.2 continuação da página anterior.

CATEGORIA	DIMENSÃO	CONCEITO	BOA PRÁTICA	
			Nº	Descrição
Confiabilidade	Grau em que as informações são aceitas como confiáveis		4	Fornecer informações sobre a licença de dados
			5	Fornecer informações sobre a proveniência dos dados
			6	Fornecer informação sobre a qualidade dos dados
			7	Fornecer indicador de versões
			8	Fornecer histórico de versões
			11	Atribuir URIs para versões de <i>datasets</i> e séries
			15	Reusar vocabulários, de preferência os padronizados
			22	Fornecer explicação para dados que não estejam disponíveis
			25	Fornecer documentação completa para a API
			26	Evitar mudanças que possam afetar o API
			27	Preservar os identificadores
			28	Avaliar a cobertura do conjunto de dados
			29	Coletar <i>feedback</i> dos consumidores dos dados
			30	Tornar o <i>feedback</i> disponível
			31	Enriquecer dados por meio da geração de novos dados
			32	Fornecer representações complementares dos dados
			33	Fornecer <i>feedback</i> para o publicador original
			34	Seguir os termos das licenças
			35	Citar a publicação original
Compreensibilidade	Entendimento dos dados sem ambiguidade		15	Reusar vocabulários, de preferência os padronizados
			25	Fornecer documentação completa para a API

Tabela 4.2 continuação da página anterior.

CATEGORIA	DIMENSÃO	CONCEITO	BOA PRÁTICA	
			Nº	Descrição
	Pontualidade	Dados atualizados	21	Fornecer dados atualizados

Fonte: Autoria própria.

## 4.2 VOCABULÁRIOS PADRONIZADOS

Os vocabulários são compreendidos como conjuntos de termos que padronizam a descrição de recursos (Tomoyose 2021). A padronização de vocabulários é essencial para o desenvolvimento da Web Semântica. Ele auxilia na integração de dados e evita ambiguidade entre termos utilizados em diferentes *datasets*, podendo até gerar novos relacionamentos entre os dados (W3C 2015).

Os vocabulários são descritos por meio de URIs. Essas URIs são “pontes” que permitem a máquinas acessarem as classes e as propriedades do vocabulário (Laufer 2015). A Tabela 4.3 apresenta alguns prefixos e URIs de vocabulários recomendados pelo W3C.

Tabela 4.3: Lista de vocabulários com seus respectivos prefixos e URIs

VOCABULÁRIO	PREFIXO	URI
Dublin Core Metadata Initiative	dc	<http://purl.org/dc/elements/1.1/>
	dct	<http://purl.org/dc/terms/>
	dctype	<http://purl.org/dc/dcmitype/>
Friend of a Friend Vocabulary	foaf	<http://xmlns.com/foaf/0.1/>
Simple Knowledge Organization System	skos	<http://www.w3.org/2004/02/skos/core#>
The PROV Ontology	prov	<http://www.w3.org/ns/prov#>
Data Catalog Vocabulary	dcat	<http://www.w3.org/ns/dcat#>
RDF Vocabulary	rdf	<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

Fonte: (W3C 2020).

Os vocabulários podem ser utilizados através de seus URIs completos ou por abreviações através de prefixos, como mostram os Códigos 4.1 e 4.2.

```

1 <html xmlns="http://www.w3.org/1999/xhtml">
2   <head>
3     <title>My home-page</title>
4     <meta property="http://purl.org/dc/terms/creator" content="Mark Birbeck" />
5     <link rel="http://xmlns.com/foaf/0.1/topic" href="http://www.example.com/#us"
6       />
7   </head>
8   <body>...</body>
9 </html>
```

Código 4.1: Vocábulos escritos com URIs completos. Fonte: (Adida et al. 2015).

```

1 <html xmlns="http://www.w3.org/1999/xhtml"
2   prefix="foaf: http://xmlns.com/foaf/0.1/
3     dc: http://purl.org/dc/terms/">
4   <head>
5     <title>My home-page</title>
6     <meta property="dc:creator" content="Mark Birbeck" />
7     <link rel="foaf:topic" href="http://www.example.com/#us" />
8   </head>
9   <body>...</body>
10 </html>

```

Código 4.2: Vocábulo escritos com prefixos. Fonte: (Adida et al. 2015).

A seguir serão apresentados os vocabulários mais utilizados por publicadores:

#### 4.2.1 Dublin Core Metadata Initiative (DC)

O Dublin Core Metadata Initiative (DCMI) teve sua origem em 1995, sendo um dos primeiros movimentos a trabalhar com a criação de um vocabulário para padronizar metadados (Laufer 2015). Este vocabulário foi criado com o propósito de gerar metadados que permitissem a recuperação de recursos informacionais eletrônicos. Atualmente o vocabulário tem por finalidade a criação de metadados que promovam a interoperabilidade com base nos princípios da Web Semântica (Dublin Core Metadata Initiative 2021, Tomoyose 2021). O DCMI possui 4 namespaces (Dublin Core Metadata Initiative 2021):

- **<http://purl.org/dc/elements/1.1/> (/elements/1.1):** corresponde ao escopo original do vocabulário. Foi criado em 2000. Lançado com 15 elementos, vem sendo amplamente utilizado por mais de vinte anos. Foi revisado no ano de 2017;
- **<http://purl.org/dc/terms/> (/terms/):** Desenvolvido em 2001, esse namespace teve como propósito agregar termos que não foram cunhados no Dublin Core original. Em 2008, os 15 termos originais foram inseridos na versão mais recente com o objetivo de dar suporte a aplicativos RDF. Pode-se tratar os termos como equivalentes, embora o DCMI incentive o uso do /terms/namespace;
- **<http://purl.org/dc/dcmitype/> (/dcmitype/):** Criado em 2001 para definir classes de tipos básicos de recursos que podem ser descritos usando termos DCMI;
- **<http://purl.org/dc/dcam/> (/dcam/):** Originou-se em 2008 com termos utilizados na descrição dos termos de metadados DCMI.

A Tabela 4.4 mostra os termos relacionados a cada namespace.

O Código 4.3 apresenta um exemplo da aplicação do vocabulário Dublin Core utilizando a tag <meta> de HTML.

Tabela 4.4: Namespaces do Vocabulário Dublin Core.

NAMESPACE	TERMO
Propriedades no namespace /terms/	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 namespace /elements/1.1/	contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type
Vocabulário do tipo DCMI (DCMI Type)	Collection, Dataset, Event, Image, InteractiveResource, MovingImage, PhysicalObject, Service, Software, Sound, StillImage, Text
Termos para descrição de vocabulário	domainIncludes , memberOf , rangeIncludes, VocabularyEncodingScheme

Fonte: (Dublin Core Metadata Initiative 2021). Adaptado.

```

1 <head>
2 <link rel="schema:DC" href="http://purl.org/dc/elements/1.1/">
3 <meta name="DC.title" content="Guia de Web Semântica">
4 <meta name="DC.creator" content="Carlos Laufer">
5 <meta name="DC.subject" content="Web Semântica">
6 <meta name="DC.subject" content="Dados Conectados">
7 <meta name="DC.description" content="Introdução ao Ecossistema e às Tecnologias de
  Web Semântica e Dados Conectados">
8 <meta name="DC.publisher" content="W3C Brasil">
9 <meta name="DC.date" content="21/01/2015">
10 </head>

```

Código 4.3: Exemplo de código com vocabulário Dublin Core. Fonte: (Laufer 2015).

## 4.2.2 Friend of a Friend Vocabulary (FOAF)

“FOAF é um projeto dedicado a vincular pessoas e informações através da web” (Brickley e Mille 2014). No vocabulário FOAF, as recursos (coisas) são chamadas de Classes e as Propriedades são informações sobre essas classes que são dadas através de links. Por exemplo, a classe `foaf:Person` pode ser vinculada à propriedade `foaf:nome` ou `foaf:homepage`, conforme exemplo do Código 4.4.

```

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

```

Código 4.4: Exemplo de código com vocabulário FOAF. Fonte: (Brickley e Mille 2014).

A Tabela 4.5 apresenta os termos das Classes e Propriedades do Vocabulário FOAF.

Tabela 4.5: Classes e Propriedades do Vocabulário FOAF.

CATEGORIA	TERMOS
Classes	Agent, Documen, Group, Image, LabelProperty, OnlineAccount, OnlineChatAccount, OnlineEcommerceAccount, OnlineGamingAccount, Organization, Person, PersonalProfileDocument, Project.
Propriedades	account, accountName, accountServiceHomepage, age, aimChatID, based_near, birthday, currentProject, depiction, depicts, dnaChecksum, familyName, family_name, firstName, focus, fundedBy, geekcode, gender, givenName, givenname, holdsAccount, homepage, icqChatID, img, interest, isPrimaryTopicOf, jabberID, knows, lastName, logo, made, maker, mbox, mbox_sha1sum, member, membershipClass, msnChatID, myersBriggs, name, nick, openid, page, pastProject, phone, plan, primaryTopic, publications, schoolHomepage, sha1, skypeID, status, surname, theme, thumbnail, tipjar, title, topic, topic_interest, weblog, workInfoHomepage, workplaceHomepage, yahooChatID

Fonte: Adaptado de (Brickley e Mille 2014).

O arquivos FOAF são documentos de textos que adotam conversações Resource Description Framework (RDF), podendo ser escritos RDFa, XML, N3 ou qualquer linguagem RDF.

### 4.2.3 Simple Knowledge Organization System (SKOS)

Simple Knowledge Organization System (SKOS) é uma recomendação W3C que contém especificações e padrões para representar e organizar o conhecimento através de tesouros e taxonomias, por exemplo. No SKOS, os recursos ou conceitos são identificados a partir de URIs e estes podem se relacionar de forma hierárquica, através de redes informais ou grupos (W3C 2012, Laufer 2015). O uso do modelo RDF permite que o SKOS seja utilizado com metadados descentralizados e distribuídos (W3C 2012).

Os principais elementos do SKOS são Concept, prefLabel/altLabel, broader/narrower, related, definition e ConceptScheme (Laufer 2015). No exemplo apresentado no Código 4.5 há três elementos: skos:Concept, skos:prefLabel e skos:related. O elemento skos:Concept diz respeito aos recursos *ornithology* e *bird* e o skos:prefLabel é um rótulo que permite identificar esses recursos. Já o elemento skos:related está relacionado os conceitos.



```

1 ex:ornithology rdf:type skos:Concept ;
2   skos:prefLabel "ornithology" .
3 ex:birds rdf:type skos:Concept ;
4   skos:prefLabel "birds" ;
5   skos:related ex:ornithology

```

Código 4.5: Exemplo de código com vocabulário SKOS. Fonte: (Laufer 2015).

#### 4.2.4 The PROV Ontology (PROV-O)

O vocabulário PROV-O é uma ontologia dedicada a descrever a procedência de recursos. Esse vocabulário tem inúmeras possibilidades de aplicações. As classes e propriedades podem ser usadas não apenas para promover informações de referência, mas também para dar detalhes específicos sobre a origem dos recursos. Dessa forma, o vocabulário pode ser utilizado para facilitar a geração de dados com proveniência interoperável (Gil e Miles 2013).

O PROV-O possui três elementos básicos: entidades, atividades e agentes. Esses elementos se relacionam de forma a estabelecer a procedência dos recursos, conforme representado na Figura 4.2. No esquema, um recurso foi gerado por um agente, que foi o responsável por realizar a atividade (Laufer 2015).

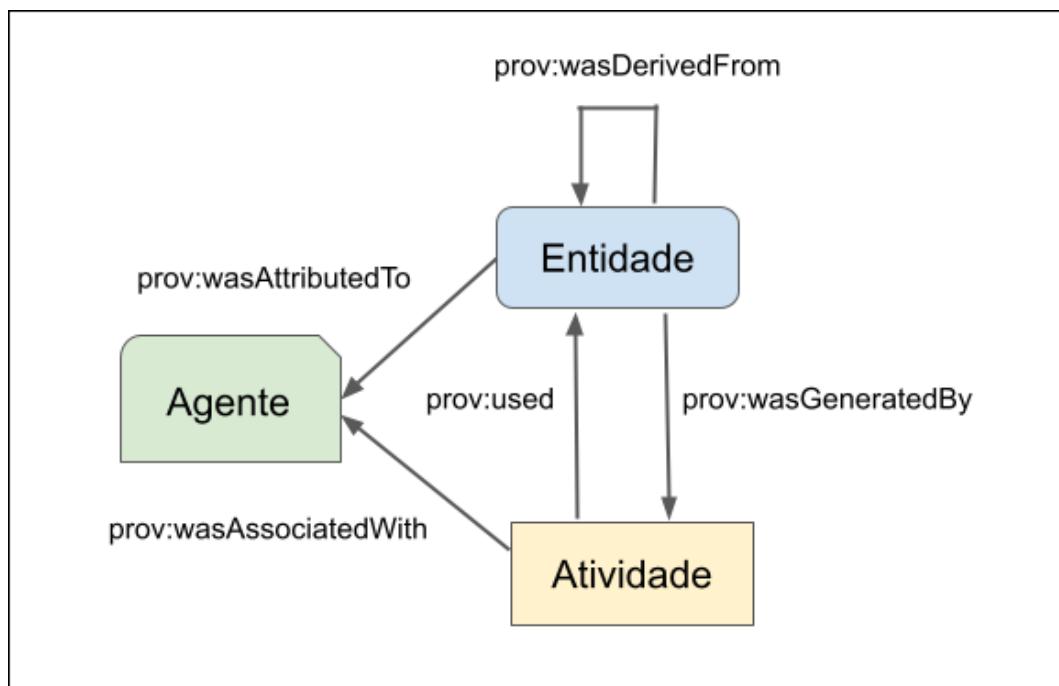


Figura 4.2: Elementos básicos do vocabulário PROV-O: entidade, atividade e agente.

Fonte: Adaptado de (Gil e Miles 2013).

Para que os usuários usem o PROV-O de acordo com suas necessidades, os termos foram divididos em 3 categorias: termos do ponto de partida, termos expandidos e termos para relacionamentos de qualificação. Os termos de pontos de partida descrevem procedências simples, já os termos expandidos e termos para relacionamentos de qualificação são mais específicos. A Tabela 4.6 apresenta os termos das classificações.

Tabela 4.6: Classificação, classes e propriedades do vocabulário PROV-O

<b>CLASSIFICAÇÃO</b>	<b>CLASSES</b>	<b>PROPRIEDADES</b>
Termos do ponto de partida	Entity, Activity, Agent	wasGeneratedBy, wasDerivedFrom, wasAttributedTo, startedAtTime, used, wasInformedBy, endedAtTime, wasAssociatedWith
Termos expandidos	actedOnBehalfOf, Collection, EmptyCollection, Bundle, Person, Location SoftwareAgen, Organization	alternateOf, specializationOf, generatedAtTime, hadPrimarySource, value, wasQuotedFrom, wasRevisionOf, invalidatedAtTime, wasInvalidatedBy, hadMembe, wasStartedBy, wasEndedBy, invalidated, influenced, atLocation, generated
Termos expandidos	Influence, EntityInfluence, Usage, Start, End, Derivation, PrimarySource, Quotation, Revision, ActivityInfluence, Generation, Communication, Invalidation, AgentInfluence, Attribution, Association, Plan, Delegation, Role InstantaneousEvent,	wasInfluencedBy, CertifiedInfluence, qualificadaGeração, qualificadaDerivação, qualificadaPrimarySource, CertifiedQuotation, qualificadaRevisão, qualificadaAtribuição, qualificadaInvalidação, qualificadaStart, qualificadaUsage, comunicação, qualificada, ficadaAssociação, qualificadaEnd, qualificadaDelegação, influenciador, entidade, hadUsage, hadGeneration, atividade, agente, hadPlan, hadActivity, atTime, hadRole

Fonte: Adaptado de (Gil e Miles 2013).

Como exemplo, o Código 4.6 retrata a história de Derek (o agente), funcionário da empresa *Chart Generations Inc*, que registrou a procedência de um conjunto de dados e de uma lista de regiões e gerou uma ilustração com esses conjuntos de dados.

```

1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix prov: <http://www.w3.org/ns/prov#> .
3 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
4 @prefix exc: <http://examplec.com/> .
5 @prefix exg: <http://exampleg.com/> .
6 exg:dataset1 a prov:Entity .
7 exc:regionList a prov:Entity .
8 exc:composition1 a prov:Entity .
9 exc:chart1 a prov:Entity .
10 exc:compose1 a prov:Activity .
11 exc:illustrate1 a prov:Activity .
12 exc:compose1 prov:used exg:dataset1 ;
13     prov:used exc:regionList1 .
14 exc:composition1 prov:wasGeneratedBy exc:compose1 .
15 exc:illustrate1 prov:used exc:composition1 .
16 exc:chart1 prov:wasGeneratedBy exc:illustrate1 .
17 exc:compose1 prov:wasAssociatedWith exc:derek .
18 exc:illustrate1 prov:wasAssociatedWith exc:derek .
19 exc:chart1 prov:wasAttributedTo exc:derek .
20 exc:derek a prov:Agent ,
21     prov:Person ;
22     foaf:givenName "Derek"^^xsd:string ;
23     foaf:mbox <mailto:derek@example.org> .
24 exc:chartgen a prov:Agent ,
25     a prov:Organization ;
26     foaf:name "Chart Generators Inc" .
27 exc:derek prov:actedOnBehalfOf exc:chartgen

```

Código 4.6: Aplicação de vocabulário PROV-O (Turtle). Fonte: (Farias et al. 2015).

#### 4.2.5 Data Catalog Vocabulary (DCAT)

DCAT é um vocabulário que segue padrões RDF cujo objetivo é facilitar a interoperabilidade entre catálogos publicados na Web (Albertoni et al. 2020). Esse vocabulário é uma ontologia OWL2 formalizada utilizando RDF Schema, sendo o OWL2 Web Ontology Language “uma linguagem desenvolvida para o contexto da Web Semântica”(Tomoyose 2021). Este vocabulário utiliza de vários outros vocabulários externos, em especial o DCTERMS, FOAF e PROV-O, podendo ser apresentado em vários formatos, tanto incorporadas a páginas Web como o HTML-RDFa, como em arquivos serializados como RDF/XML, N3, Turle ou JSON-DL. O vocabulário conta com 6 classes principais (Albertoni et al. 2020):

- **dcatalog:Catalog:** A classe `dcatalog:Catalog` é uma coleção de metadados acerca de conjuntos de dados ou serviços de dados, onde cada metadado descreve um recurso;
- **dcatalog:Resource:** Caracteriza um conjunto de dados, serviços de dados ou outro recurso representado por metadados de um catálogo. Essa classe não é utilizada diretamente nos arquivos de dados. Ela é uma classe “classe pai” de `dcatalog:Dataset`, `dcatalog:DataService` e `dcatalog:Catalog`;

- **dcat:Dataset**: Dataset é a classe que representa um conjunto de dados, ou seja, uma coleção de dados “publicados ou com curadoria de um único agente ou comunidade identificável”;
- **dcat:Distribution**: Diz respeito ao como o conjunto de dados pode ser acessado como, por exemplo, arquivo para download;
- **dcat:DataService**: Corresponde a um serviço de dados, que uma manipulação de um ou mais conjuntos de dados feita através de uma interface (API);
- **dcat:CatalogRecord**: Caracteriza um informações sobre dados em um catálogo, por exemplo, a hora de criação do dado ou sua origem.

Tabela 4.7: Classes e propriedades do DCAT.

CLASSE	PROPRIIDADE
dcat:Catalog	foaf:homepage, dcat:themeTaxonomy, dct:hasPart, dcat:dataset, dcat:service, dcat:catalog, dcat:record
dcat:Resource	dct:accessRights, dct:conformsTo, dcat:contactPoint, dct:creator, dct:description, dct:title, dct:issued, dct:modified, dct:language, dct:Publisher, dct:identifier, dcat:theme, dct:type, dct:relation, dcat:qualifiedRelation, dcat:keyword, dcat:landingPage, prov:qualifiedAttribution, dct:license, dct:rights, odrl:hasPolicy, dct:isReferencedBy.
dcat:Record	dct:title, dct:description, dct:issued, dct:modified, foaf:primaryTopic, dct:conformsTo
dcat:Dataset	dcat:distribution, dct:accrualPeriodicity, dct:spatial, dc: temporal, dcat:spatialResolutionInMeters, dcat:temporalResolution, prov:wasGeneratedBy
dcat:Distribution	dct:title, dct:description, dct:issued, dct:modified, dct:license, dct:accessRights, dct:rights, odrl:hasPolicy, dcat:accessURL, dcat:accessService, dcat:downloadURL, dcat:byteSize, dcat:spatialResolutionInMeters, dcat:temporalResolution, dct:conformsTo, dcat:mediaType
dcat:Service	dcat:endpointURL, dcat:endpointDescription, dcat:servesDataset
skos:Concept	-
foaf:Organization/Person	-
dcat:Relationship	dct:relation, dcat:hadRole
dcat:Role	-
dct:PeriodOfTime	dcat:startDate, dcat:endDate, time:hasBeginning, time:hasEnd
dct:Location	locn:geometry, dcat:bbox, dcat:centroid

Fonte: (Albertoni et al. 2020)

Pode-se perceber que algumas propriedades se repetem dentro de classes diferentes. Por exemplo, a propriedade `dct:title` está presente em três classes: `cat:Resource`, `dcat:CatalogRecord` e `dcat:Distribution`. Isso acontece porque cada atributo tem uma semântica diferente, sendo que o `dcat:Resource` se refere ao nome do item, `dcat:CatalogRecord` está relacionando ao nome do registro e o `dcat:Distribution`, ao nome da distribuição (Tomoyose 2021).

Na Figura 4.3 pode ser visualizado um diagrama que mostra as relações entre as classes do DCAT. O Código 4.7 mostra a aplicação de alguns atributos nesse vocabulário.

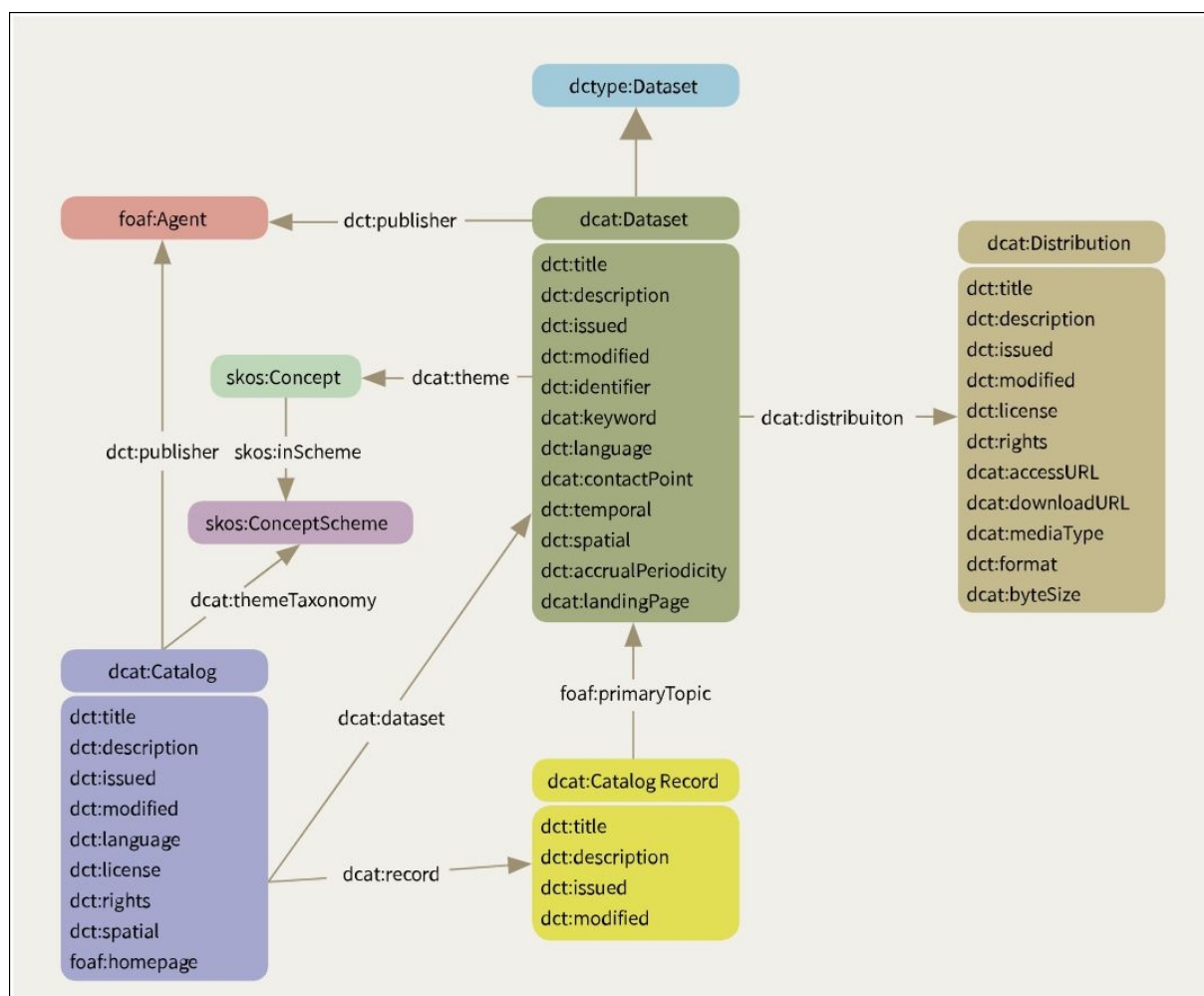


Figura 4.3: Diagrama de Classes de DCAT.

Fonte: (Laufer 2015)

```

1 : catalog
2   a dcat:Catalog ;
3   dct:title "Imaginary Catalog"@en ;
4   rdfs:label "Imaginary Catalog"@en ;
5   foaf:homepage <http://example.org/catalog> ;
6   dct:publisher :transparency-office ;
7   dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
8   dcat:dataset :dataset-001 , :dataset-002 , :dataset-003 ;

```

Código 4.7: Aplicações de atributos do vocabulário DCAT. Fonte: (Albertoni et al. 2020).

### 4.3 REPRESENTAÇÕES SINTÁTICAS PARA O MODELO RDF 1.1

Segundo (Jevsikova et al. 2017), o RDF é utilizado para metadados em informações semânticas que apresentam relações contextuais. Isso corrobora com a interoperabilidade, facilidade de buscar e consequente recuperação de objetos.

Devido ao fato de RDF ser um modelo abstrato, a forma com que ele é representado não interfere no processamento dos dados, desde que a representação siga as propriedades abstratas do modelo. Há várias maneiras como o RDF pode ser representado ou serializado (Isotani e Bittencourt 2015). Não existe um modelo de serialização melhor que outro. Desenvolvedores costumam escolher de acordo com suas necessidades como, por exemplo, velocidade de processamento ou facilidade de leitura humana (Isotani e Bittencourt 2015, Lóscio, Burle e Calegari 2017).

Os modelos de serialização do RDF apresentaram um acréscimo considerável da versão RDF1 para a RFD 1.1, conforme mostra a Figura 4.4.

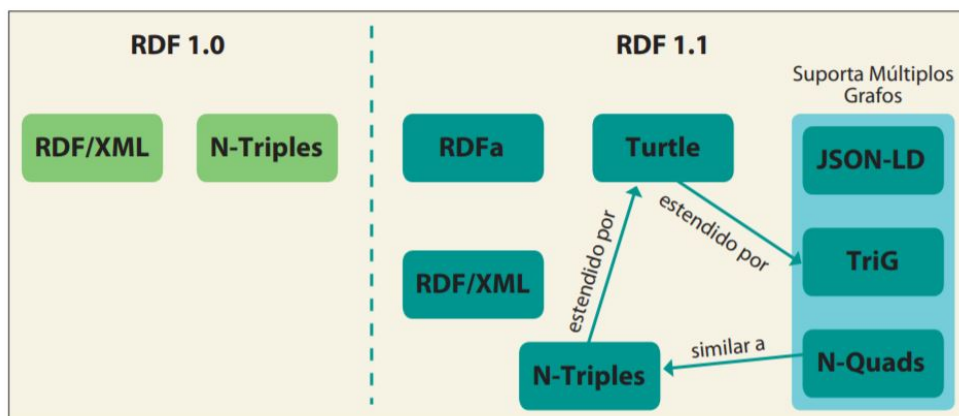


Figura 4.4: Formatos de serialização das versões RDF1 e RDF1.1.

Fonte: (Isotani e Bittencourt 2015). Adaptado de (Wood et al. 2014)

Os modelos RDF1.1. são apresentados nas seguintes subseções.

#### 4.3.1 RDF/XML

RDF/XML é considerada a primeira serialização padronizada RDF. Esta representação segue a sintaxe de grafos do RDF (Isotani e Bittencourt 2015). De acordo com (Farias et al. 2015), “seu uso inicial tinha como vantagem o fato de as linguagens de programação terem mais suporte para XML. Além disso, também é possível fazer uso de namespaces XML para evitar o uso de URIs completas, o que torna as URIs menos extensas”. A desvantagem da representação está na dificuldade de interpretação humana, o que torna o seu uso mais complexo.

O Código 4.8 representa as informações da Tabela 4.8. Pode-se notar que toda a estrutura do código está dentro da tag `<rdf:RDF>`. O `rdf:Description` é um elemento utilizado para identificar o sujeito, enquanto o `rdf:about` o descreve. Essa é uma característica importante no RDF/XML (Isotani e Bittencourt 2015).

Tabela 4.8: Dados utilizados para descrição do Código RDF/XML.

Título	Artista	País	Empresa	Preço	Ano
Empire Burlesque	Bob Dylan	USA	Columbia	10.90	1985
Hide your heart	Bonnie Tyler	UK	CBS Records	9.90	1988

Fonte: Adaptado de (W3Schools 2021).

```

1 <?xml version="1.0"?>
2 <rdf:RDF
3   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
4   xmlns:cd="http://www.recshop.fake/cd#">
5   <rdf:Description
6     rdf:about="http://www.recshop.fake/cd/Empire_Burlesque">
7     <cd:artist>Bob Dylan</cd:artist>
8     <cd:country>USA</cd:country>
9     <cd:company>Columbia</cd:company>
10    <cd:price>10.90</cd:price>
11    <cd:year>1985</cd:year>
12  </rdf:Description>
13
14  <rdf:Description
15    rdf:about="http://www.recshop.fake/cd/Hide_your_heart">
16    <cd:artist>Bonnie Tyler</cd:artist>
17    <cd:country>UK</cd:country>
18    <cd:company>CBS Records</cd:company>
19    <cd:price>9.90</cd:price>
20    <cd:year>1988</cd:year>
21  </rdf:Description>
22 </rdf:RDF>

```

Código 4.8: Exemplo de código na serialização RDF/XML. Fonte: (W3Schools 2021).

### 4.3.2 RDFa

O RDFa é uma representação RDF baseada em atributos, que podem ser embutidos dentro de um código HTML ou XML (Herman et al. 2015). Os atribuídos embutidos nas páginas aumentam a precisão das máquinas durante as buscas na Web e possibilitam o enriquecimento da pesquisa ao permitir mesclar dados de diferentes documentos (Isotani e Bittencourt 2015).

No Código 4.9, pode-se visualizar três atributos: `property`, `typeof` e `href`. O atributo `property` representa a relação entre o sujeito e o objeto da tripla RDF, `typeof` é “equivalente ao atributo para representar o elemento `rdf:type`” e `href` é usado para navegação entre URIs (Isotani e Bittencourt 2015,

Adida et al. 2015).

```
1 <div vocab="http://schema.org/" typeof="Person">
2   <a property="image" href="http://manu.sporny.org/images/manu.png">
3     <span property="name">Manu Sporny</span></a>,
4   <span property="jobTitle">Founder/CEO</span>
5   <div>
6     Phone: <span property="telephone">(540) 961-4469</span>
7   </div>
8   <div>
9     E-mail: <a property="email" href="mailto:msporny@digitalbazaar.com">
10      msporny@digitalbazaar.com</a>
11   </div>
12   <div>
13     Links: <a property="url" href="http://manu.sporny.org/">Manu's homepage</a>
14   </div>
15 </div>
```

Código 4.9: Exemplo de código com serialização RDFa. Fonte: (Herman et al. 2015).

### 4.3.3 JSON-LD

O JSON-LD é um formato de serialização que tem como objetivo expandir o JSON para a representação de Dados Conectados. Ele foi projetado para que programadores possam passar tranquilamente de JSON para JSON-DL. É usada principalmente para construção de serviços Web e interoperáveis em sistemas JSON (Sporny, et al. 2020). No Código 4.10, o `@context` informa a máquina as ferramentas necessárias para entender os termos usados durante o código e evitar ambiguidades. Faz-se isso através da determinação de vocabulários e outras definições relevantes. O termo `@id` é interpretado por uma URI.

```
1 {
2   "@context": {
3     "name": "http://schema.org/name",
4     "image": {
5       "@id": "http://schema.org/image",
6       "@type": "@id"
7     },
8     "homepage": {
9       "@id": "http://schema.org/url",
10      "@type": "@id"
11    }
12  },
13  "name": "Manu Sporny",
14  "homepage": "http://manu.sporny.org/",
15  "image": "http://manu.sporny.org/images/manu.png"
16 }
```

Código 4.10: Exemplo de código com serialização JSON-LD. Fonte: (Sporny, et al. 2020).



### 4.3.4 N-Triples

N-Triples é a serialização mais simples de todas por se tratar de uma declaração direta da tripla sujeito, predicado e objetivo do RDF. Nessa serialização, cada bloco de linhas representa um tripla. Por exemplo, no Código 4.11, as primeiras 3 linhas estão relacionadas de modo de Bob <http://example.org/bob#me> é um tipo <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> de pessoa <http://xmlns.com/foaf/0.1/Person>.

```
1 01 <http://example.org/bob#me>
2   <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
3   <http://xmlns.com/foaf/0.1/Person> .
4 02 <http://example.org/bob#me>
5   <http://xmlns.com/foaf/0.1/knows>
6   <http://example.org/alice#me> .
7 03 <http://example.org/bob#me>
8   <http://schema.org/birthdate> "1990-07-04"^^
9   <http://www.w3.org/2001/XMLSchema#date> .
10 04 <http://example.org/bob#me>
11   <http://xmlns.com/foaf/0.1/topic_interest>
12   <http://www.wikidata.org/entity/Q12418> .
13 05 <http://www.wikidata.org/entity/Q12418>
14   <http://purl.org/dc/terms/title> "Mona Lisa" .
15 06 <http://www.wikidata.org/entity/Q12418>
16   <http://purl.org/dc/terms/creator>
17   <http://dbpedia.org/resource/Leonardo_da_Vinci> .
18 07 <http://data.europeana.eu/item/04802/243FA8618938F419AA4D619>
19   <http://purl.org/dc/terms/subject>
```

Código 4.11: Exemplo de código com serialização N-Triples. Fonte: Adaptado de (Isotani e Bittencourt 2015).

### 4.3.5 N-Quads

A serialização N-Quads é um expansão do N-Triples, sendo utilizada para trocas entre catálogos de dados (Isotani e Bittencourt 2015), como pode ser visualizado no Código 4.12.

```
1 01 <http://example.org/bob#me>
2   <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
3   <http://xmlns.com/foaf/0.1/Person> .
4 02 <http://example.org/bob#me>
5   <http://xmlns.com/foaf/0.1/knows>
6   <http://example.org/alice#me> .
7 03 <http://example.org/bob#me>
8   <http://schema.org/birthDate> "1990-07-04"^^^^
9   <http://www.w3.org/2001/XMLSchema#date> .
10 04 <http://example.org/bob#me>
11   <http://xmlns.com/foaf/0.1/topic\_interest>
12   <http://www.wikidata.org/entity/Q12418> .
```

```

13 05 <http://www.wikidata.org/entity/Q12418>
14   <http:purl.org/dc/terms/title> "Mona Lisa" .
15 06 <http://www.wikidata.org/entity/Q12418>
16   <http:purl.org/dc/terms/creator>
17   <http://dbpedia.org/resource/Leonardo_da_Vinci> .
18 07 <http://data.europeana.eu/item/04802/A8B813C5F9AA4D619>
19   <http:purl.org/dc/terms/subject> .
20 08 <http://example.org/bob#me>
21   <http:purl.org/dc/terms/publisher>
22   <http://example.org> .
23 09 <http://example.org/bob#me>
24   <http:purl.org/dc/terms/rights>
25   <http://creativecommons.org/licences/by/3.0/> .

```

Código 4.12: Exemplo de código com serialização N-Quads. Fonte: Adaptado de (Isotani e Bittencourt 2015).

### 4.3.6 Turtle

Assim como os outros tipos de serialização RDF, o turtle segue a representação sujeito, predicado e objeto. Possui uma forma textual compacta, com sintaxe bastante simples e de fácil compreensão para humanos. Ela foi criada como uma evolução do N-Triples para permitir o uso de prefixos e URIs relativos dentro do código (Beckett e Berners-Lee 2011, Isotani e Bittencourt 2015). A seguir, o Código 4.13 descreve a relação entre o Duende Verde (Green Goblin) e o Homem-Aranha (Spiderman).

```

1 base <http://example.org/> .
2 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
3 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
4 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
5 @prefix rel: <http://www.perceive.net/schemas/relationship/> .
6
7 <#green-goblin>
8   rel:enemyOf <#spiderman> ;
9   a foaf:Person ;    # in the context of the Marvel universe
10  foaf:name "Green Goblin" .
11
12 <#spiderman>
13   rel:enemyOf <#green-goblin> ;
14   a foaf:Person ;
15   foaf:name "Spiderman" , "Homem-Aranha"@pt .

```

Código 4.13: Exemplo de código com serialização Turtle. Fonte: Adaptado de (Prud'hommeaux e Carothers 2014).

### 4.3.7 TriG

TriG é uma extensão do Turtle, por isso tem a mesma simplicidade de síntese e de leitura humana. Essa serialização foi criada para múltiplos graficos, uma atualização trazida pela versão 1.1 do RDF. A grande diferença entre o Turtle e o TriG é o uso do elemento GRAPH. Esse elemento tem como função representar

um conjunto de dados dentro do código (Carothers e Seaborne 2014, Isotani e Bittencourt 2015), como pode ser visualizado no Código 4.14.

```
1 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
2 @prefix dc: <http://purl.org/dc/terms/> .
3 @prefix foaf: <http://xmlns.com/foaf/0.1/> .
4
5 # default graph
6 {
7     <http://example.org/bob> dc:publisher "Bob" .
8     <http://example.org/alice> dc:publisher "Alice" .
9 }
10
11 <http://example.org/bob>
12 {
13     _:a foaf:name "Bob" .
14     _:a foaf:mbox <mailto:bob@oldcorp.example.org> .
15     _:a foaf:knows _:b .
16 }
17
18 <http://example.org/alice>
19 {
20     _:b foaf:name "Alice" .
21     _:b foaf:mbox <mailto:alice@work.example.org> .
22 }
```

Código 4.14: Exemplo de código com serialização TriG. Fonte: (Carothers e Seaborne 2014).

# 5 ELICITAÇÃO E APLICAÇÃO DE REQUISITOS PARA GERAÇÃO DE RECURSOS EDUCACIONAIS ABERTOS CONECTADOS

De forma geral, uma Boa Prática para a publicação de dados na Web está em estabelecer padrões. As Boas Práticas estão relacionadas ao acesso aos dados, fornecimento de metadados, uso de identificadores únicos e divulgação de arquivos de diversos formatos (Lóscio, Burle e Calegari 2017).

Este Capítulo tem como propósito a elicitação de requisitos que apoiam a criação de Recursos Educacionais Abertos Conectados, assim como a apresentação de uma proposta de representação semântica para projetos educacionais de acordo com as recomendações do W3C. Como exemplo, é utilizado uma sugestão de módulo para Projeto PUMA: Cadastro de Resultados. Esse projeto é desenvolvido pela Universidade de Brasília e tem por objetivo a documentação e o acompanhamento dos resultados obtidos durante a aplicação da metodologia Aprendizagem Baseada em Projetos (do inglês, *Project Based Learning* - PBL) no Curso de Engenharia de Produção.

## 5.1 ELICITAÇÃO DE REQUISITOS

De acordo com (Bandeira et al. 2015), para produzir dados conectados, deve-se aplicar um conjunto de requisitos de qualidade. Alguns desses requisitos são obtidos através da aplicação das boas práticas para a publicação de dados conectados. Segundo os autores, ainda, para se estabelecer “vocabulários e ontologias, são necessárias ferramentas de engenharia de ontologias ou ferramentas de organização de palavras-chave”.

A Figura 5.1 apresenta os Princípios para Dados Abertos, as Melhores Práticas para Publicação de Dados Conectados e as Boas Práticas para Dados na Web. Com base nesses fundamentos, internacionalmente conceituados, foram propostos os requisitos para a geração Recursos Educacionais Abertos Conectados. Esses requisitos foram pensados de forma a considerar fatores importantes no momento da criação do banco de dados de projetos educacionais. As categorias consideradas foram: Metadados, Formatos de dados, Uso de vocabulários padrões, Licença de dados, Identificadores de dados, Coleta e Compartilhamento de *Feedbacks*.

A Tabela 5.1 traz as categorias, os requisitos e os princípios e/ou melhores práticas em que esses requisitos estão baseados.

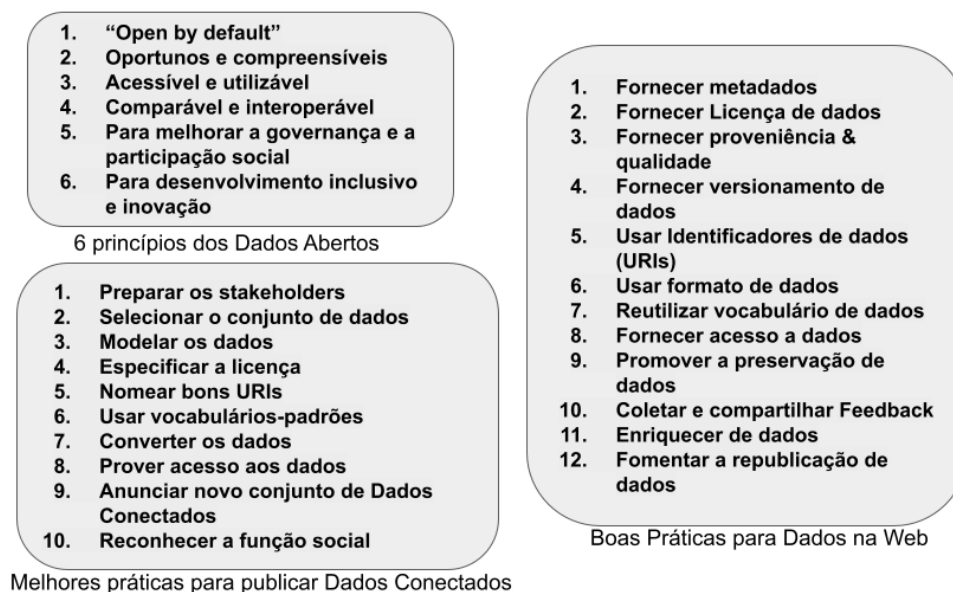


Figura 5.1: Princípios e melhores praticas para Dados Abertos, Dados Conectados e Dados na Web.

Fonte: Autoria própria, 2021.

Tabela 5.1: Requisitos para a geração de Recursos Educacionais Abertos Conectados.

<b>CATEGORIA</b>	<b>REQUISITO</b>	<b>FONTE</b>
Metadados	Fornecer metadados básicos	1ª Boa prática para dados na web 7ª Melhor Prática para Publicação de Dados Conectados
	Fornecer metadados descritivos	2ª Boa prática para dados na web
Formatos de dados	Usar formatos não proprietários	3º Principio dos dados abertos
	Usar formatos processáveis por máquinas	5º Principio dos dados abertos 12ª Boa prática para dados na web
	Fornecer dados em múltiplos formatos	14ª Boa prática para dados na web
Uso de vocabulários padrões	Utilizar vocabulários padronizados	6ª Melhor Prática para Publicação de Dados Conectados 7ª Boa prática para dados na web
	Fornecer de licença abertas	2ª Boa prática para dados na web 3º Principio dos dados abertos
Identificadores de dados	Nomear bons URIs	5ª Boa prática para dados na web 5ª Melhor Prática para Publicação de Dados Conectados
Coleta e Compartilhamento de Feedbacks	Coletar Feedbacks	29ª Boa prática para dados na web

**Tabela 5.1 continuação da página anterior.**

<b>CATEGORIA</b>	<b>REQUISITO</b>	<b>FONTE</b>
	Tornar o feedback disponível	30ª Boa prática para dados na web

Fonte: Autoria própria.

As categorias dos requisitos são descritas a seguir:

### 5.1.1 Metadados

O uso de metadados corrobora para encontrar e reutilizar dados na Web. Eles proporcionam informações adicionais para que os consumidores possam entender melhor o conceito e a estrutura do dado. Por conseguinte, os metadados auxiliam que os conjuntos de dados sejam utilizados por um público variado e não apenas pelo próprio publicador. Assim, “fornecer metadados é um requisito fundamental na publicação de dados na Web” (Lóscio, Burle e Calegari 2017).

São possíveis abordagens para implementação de metadados (Lóscio, Burle e Calegari 2017):

- **Abordagens legíveis por pessoas:** pode-se fornecer metadados embutidos em uma página Web ou em um arquivo separado;
- **Abordagens legíveis por máquinas:** Há duas formas de fornecer dados legíveis para máquinas: utilizando formatos serializados como Turtle ou JSON ou agregados esses dados nas páginas HTML, com HTML-RDFA ou JSON-LD.

De acordo com a 7ª Melhor Prática para Publicação de Dados Conectados, Converter dados para Dados Conectados, é relevante incluir metadados como organização e/ou agência de publicação, data de criação, versão, frequência de atualizações e endereço de e-mail de contato. Consoante com esta recomendação, a 2ª Boa Prática para Dados na Web orienta a utilização de metadados descritivos que possibilitem a descoberta e identificação do dado. A Figura 5.2 mostra alguns exemplos de métodos descritivos sugeridos pelo W3C.

- O **título** e a **descrição** do conjunto de dados.
- As **palavras-chave** que descrevem o conjunto de dados.
- A **data de publicação** do conjunto de dados.
- A **entidade responsável (publicadora de dados)** por disponibilizar o conjunto de dados.
- O **ponto de contato** para o conjunto de dados.
- A **cobertura espacial** do conjunto de dados.
- O **período temporal** coberto pelo conjunto de dados.
- A **data da última modificação** do conjunto de dados.
- Os **temas/categorias** cobertos por um conjunto de dados.

Figura 5.2: Exemplos de metadados descritivos.

Fonte: (Lóscio, Burle e Calegari 2017).

O W3C recomenda que a versão legível por máquinas dos metadados seja escrita com vocabulário sugerido pela instituição. Um exemplo é o Data Catalog Vocabulary (VOCAB-DCAT).

### 5.1.2 Formatos de dados

De acordo com (Lóscio, Burle e Calegari 2017), “o Formato de Arquivo é uma maneira padronizada por meio da qual a informação é codificada para armazenamento em um arquivo de computador. Ele especifica como os bits são usados para codificar informação em um meio digital de armazenamento”. Os formatos de arquivos podem ser proprietários ou livres (não proprietários), novos ou disponíveis.

Ao se tratar de Dados Conectados, deve-se ter em conta que não se pode falar nesse conceito sem fazer uso do padrão RDF. Dessa forma, levando em consideração o sistema 5 estrelas proposto por Tim Berners-Lee, somente há Dados Conectados quando a abertura dos dados contempla pelo menos 4 estrelas (Isotani e Bittencourt 2015). A quarta estrela está relacionada com o uso de URIs, que identificam recursos na Web.

(Alcantara et al. 2015) afirmam que o ideal é que os dados sejam legíveis facilmente tanto por aqueles que buscam interpretar a informação para uso próprio, sem a necessidade de manipular os dados e gerar novas aplicações (homens), quanto por aqueles que desejam realizar o processamento em massa desses dados (máquinas). Para o primeiro grupo de usuários, os formatos mais indicados são PDF, DOC, ODT, XLS, por exemplo. Enquanto o segundo grupo de usuários necessitam de formatos tais como JSON, XML e RDF.

É importante ressaltar que, para uso humano, a publicação de dados em formatos não estruturados como, por exemplo, PDF, TXT, DOC e ODT dificultam a extração dos dados (Wood et al. 2014). Por isso, deve-se priorizar o uso de dados estruturados que possibilitem a aplicação de filtros e extração de dados, assim como CSV. De fato, quanto maior a abertura de dados em relação ao formato 5 estrelas, maiores são as aplicações, conexões e interoperabilidade dos dados.

### 5.1.3 Uso de vocabulários padrões

Os vocabulários são utilizados para representar uma certa área do conhecimento. São termos próximos usados para esta finalidade: ontologia, vocabulário controlado, glossário, taxonomia, lista de códigos e rede semântica.

É através das classes e propriedade dos vocabulários que as máquinas conseguem entender os termos e criar contexto para os documentos analisados. O uso de vocabulários padronizados é fundamental para que um agente *software* consiga interpretar o conteúdo disponibilizado. Se um publicar lança na Web dados com vocabulário desconhecidos, tornando mais complicado a interpretação e conexão com outras fontes de dados.

Recomenda-se utilizar vocabulários reconhecidos e utilizados globalmente, como por exemplo, os documentos pelo W3C.

### 5.1.4 Licença de dados

A disponibilização de licenças é crucial para o dado ser considerado aberto. A licença que vai determinar para homens e máquinas qual o grau de abertura do documento. Através das licenças, os agente de *software* podem mesclar os dados com outros conjuntos de dados na Web.

Há várias formas de propiciar licença aos usuário do dataset, podendo ser “por meio de um link ou de uma cópia embutida dos termos da licença que seja legível por pessoas. Também podem ser disponibilizadas para processamento um link ou cópia embutida dos termos da licença legível por máquina” (Lóscio, Burle e Calegari 2017).

Pode-se incluir licenças através dos seguintes vocabulários: Dublin Core (DCTERMS), Creative Commons (CCREL), schema.org (SCHEMA-ORG) ou XHTML.

### 5.1.5 Identificadores de dados

O uso de identificadores únicos é o centro dos Dados Conectados. Através destes identificadores, os recursos podem ser encontrados na rede. Se os dados estão na Web, através do uso dos URIs HTTP, é muito mais fácil encontrá-los e consultá-los.

Para usufruir dos benefícios advindos do uso das URIs, é necessário pensar na criação e manutenção do identificar desde de momento da criação do conjunto de dados. Segundo as recomendações do W3C, o uso de links curtos facilita o uso do URIs. Outro fator importante é a manutenção dos identificadores. Os publicadores de dados devem estar atentos para que os URIs permaneçam ativos ao longo do tempo, Uma alternativa é o uso de como o *Permanent Identifiers for the Web*<sup>1</sup>, [purl.org](http://purl.org)<sup>2</sup>. ou DOI (Identificadores de Objetos Digitais). O uso de identificadores permanentes apoia na utilização destes em diversos conjuntos de dados.

### 5.1.6 Coleta e Compartilhamento de Feedbacks

A coleta de *feedback* promove a oportunidade que os usuários venham a contribuir para a formação e melhoria de um conjunto de dados.

Para que os publicadores garantam que os seus dados estão cumprindo com os objetivos esperados, é fundamental disponibilizar meios de contato dos consumidores com os publicadores. O *feedback* é benéfico tanto para publicadores, quanto para os consumidores, que têm a oportunidade de expressar suas impressões e necessidades. Já os publicadores ganham ferramentas para a melhoria das publicações. Sempre que possível, é uma boa prática compartilhar os *feedbacks* recebidos, de maneira estruturada, a fim de incentivar a colaboração na Web (Lóscio, Burle e Calegari 2017).

---

<sup>1</sup><https://w3id.org/>

<sup>2</sup><https://archive.org/services/purl/>



## 5.2 PROPOSTA DE REPRESENTAÇÃO SEMÂNTICA PARA REGISTRO DE RECURSOS EDUCACIONAIS

Para a aplicação dos requisitos supracitados, foi desenvolvida uma proposta de representação semântica para Recursos Educacionais Abertos Conectados em um projeto educacional. O Projeto utilizado foi a plataforma PUMA, que está descrita a seguir.

### 5.2.1 PLATAFORMA UNIFICADA DE METODOLOGIA ATIVA (PUMA)

Somente as habilidades técnicas não são mais suficientes para que os engenheiros se mantenham no mercado de trabalho. Eles também precisam desenvolver habilidades interpessoais, de comunicação e dinâmica em grupo (Deshpande e Huang 2011, Atenas et al. 2015).

Para atender ao mercado de trabalho, o curso de Engenharia de Produção da Universidade de Brasília foi projetado para promover, além das capacidades técnicas esperadas, competências transversais tais como liderança, gerenciamento e proatividade (Monteiro et al. 2017). Espera-se que o Engenheiro de Produção formado na UnB saiba “identificar, caracterizar e tratar adequadamente as formas de criação de valor associadas aos espaços econômicos, político e cultural da sociedade, levando em conta os diferentes patamares de convivência pessoal, institucional e em rede” (Silva e Balthazar 2010). Para alcançar tais objetivos, o curso utiliza o PBL, Aprendizagem Baseada em Projetos, do inglês (*Project Based Learning*).

A Aprendizagem Baseada em Projetos é uma das estratégias que integram as metodologias de aprendizagem ativa. Nessa metodologia, o professor deixa de ter o papel de guardião do conhecimento e passa a ser um facilitador (Farias et al. 2015). De acordo com (Monteiro et al. 2017) “a utilização da abordagem estimula o aluno a buscar conhecimentos, propondo soluções para problemas reais, introduzidos por agentes externos”. A aprendizagem do aluno depende da busca de conhecimentos, técnicas e ferramentas que lhe permitam resolver tais problemas. Ele torna-se, assim, o centro do processo e não mais um agente passivo.

Para aplicar o PBL, o curso possui, em sua grade curricular, sete disciplinas de projeto, os PSPs (Projetos de Sistemas de Produção) que são ministradas do quarto até o décimo semestre (Monteiro et al. 2017). Essas disciplinas são apoiadas por outras de cunho técnico, a saber: Probabilidade e Estatística, Sistema de Informação aplicado à Engenharia de Produção, Planejamento e Controle da Produção, Gestão da Qualidade, Engenharia do Produto e Gestão Estratégica (Júnior et al. 2019). A Figura 5.3 mostra o esquemático das disciplinas de PSP e suas disciplinas âncoras.

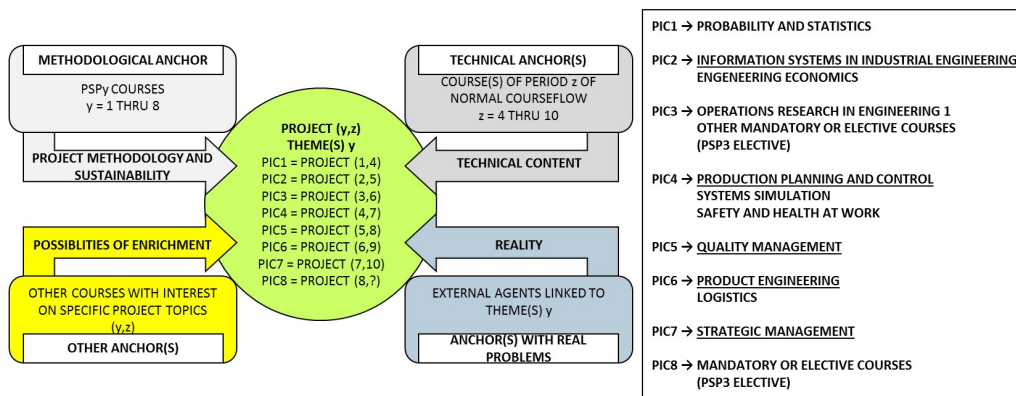


Figura 5.3: Esquema Geral para Projeto âncoras (y, z) do curso de Engenharia de Produção da UnB.

Fonte: (Monteiro et al. 2017) adapdo de (Zindel et al. 2012).

A pesquisadora (Monteiro et al. 2017) afirma que um dos grandes obstáculos da adoção da metodologia PBL é a obtenção de uma boa avaliação dos alunos ao longo da sua graduação. A fim de avaliar a evolução do aluno ao longo do curso, concebeu-se a Plataforma Plataforma Unificada de Metodologia Ativa (PUMA).

O PUMA é uma plataforma web que visa acompanhar o desenvolvimento dos estudantes durante a graduação e irá possibilitar uma melhor avaliação do processo de aprendizagem ativa por meio de *feedbacks* das partes interessadas. A plataforma visa, também, mensurar as expectativas do mercado de trabalho e alinhar com as atividades do curso com as demandas deste (A3M 2018). O projeto é apoiado pela iniciativa A3M (Aprendizagem para o 3º Milênio), que tem por objetivo “atuar junto à comunidade UnB na identificação, valorização e promoção de ações educacionais inovadoras”(A3M 2021).

Os dados que podem ser extraídos dos PSPs e dos projetos de Graduação e serem utilizados para<sup>3</sup>:

- Possibilitar o processo de filtragem de temas propostos por empresas, de acordo com as necessidades do curso e da UnB;
- Conectar alunos e professores em torno de temas de pesquisa para PIBIC ou mestrado, por exemplo;
- Propiciar uma posterior alocação e continuação das pesquisas realizadas, de forma a apoiar um maior aprofundamento dos temas estudados;
- Oportunizar a recomendação de professores para bancas de graduação, mestrado ou doutorado;
- Mensurar os temas mais trabalhados no curso de Engenharia de Produção.

No atual momento do projeto, construção do PUMA, essa base de dados está sendo criada. Entretanto é essencial que esses dados sigam padrões internacionalmente conceituados para que possam se conectar entre si e com outras bases de dados. A padronização é necessária para a geração de dados com qualidade e é essencial para a interoperabilidade destes.

<sup>3</sup>Os usos dos dados do projeto foram obtidos através das entrevistas realizadas com o PUMA.

A seguir será apresentada a aplicação de uma representação semântica em um Módulo sugerido para o Projeto PUMA.

## 5.2.2 PROPOSTA DE REPRESENTAÇÃO DE REGISTRO DE PROJETO

Na Plataforma Unificada de Metodologia Ativa estão sendo desenvolvidos os seguintes módulos: Divulgação e Submissão de Projetos, Alocação Automatizada de Projetos, Medição de Qualidade de projetos PBL pelos Stakeholders, Gestão de Ex-aluno, Avaliação da evolução dos alunos nas competências transversais e Divulgação de Menção e Resultados de disciplinas de projetos (PSPs).

Este trabalho apresenta uma proposta de Módulo para o Cadastro dos Resultados desenvolvidos nas disciplinas de Projetos de Sistemas de Produção (PSPs) e Projetos de Graduação (PGs). Nesse Módulo sugerido, os alunos deverão preencher um formulário na própria plataforma e esses dados deverão ser registrados em uma Página Web, conforme Figura 5.4. Esta figura apresenta os dados de um artigo, fruto da disciplina de PSP2, desenvolvido no primeiro semestre de 2018.

Disciplinas vinculadas	Projeto de Sistemas de Produção 2 Sistemas de Informação em Engenharia de Produção
Área/Subárea	Engenharia de Produção, Sustentabilidade e Responsabilidade Social Ética e Transparência nas Decisões Organizacionais
Título	Combate à corrupção em empresas privadas, questão de integridade e compliance: estudo de caso siemens
Palavras-chave	Compliance Integridade Corrupção Siemens
Autor(es)	Pamella Oliveira de Souza <pamella.exemplo@unb.br> Ada Maria Lopes Santos <ada.exemplo@unb.br> Salém Miranda Alves <saalem.exemplo@unb.br> Yan Vieira de Jesus Costa <yan.exemplo@unb.br>
Orientador(es)	Edgard Costa Oliveira <edgard.exemplo@unb.br>
Data de apresentação:	26 de junho de 2018
Data de publicação:	08 de abril de 2021
Resumo	A corrupção é definida como abuso público ou privado de poder para obter benefício próprio, para amigos ou familiares, está presente em todas as sociedades, de diferentes formas e áreas. Para combatê-la, depende do esforço conjunto e contínuo de todos, inclusive das empresas, assim como o controle social e estudos na área de compliance, integridade e ética. O artigo será baseado no estudo de caso sobre o sistema de compliance da empresa siemens. O presente estudo é de natureza qualitativa e utiliza o método de pesquisa descritiva a partir de fontes documentais e método indutivo com objetivo de analisar, observar, registrar e correlacionar aspectos para compreensão da gestão de compliance e integridade no combate à corrupção. A partir do estudo realizado, foi possível perceber que o sistema de compliance da siemens cumpre os requisitos da gestão de risco efetiva, pois leva em consideração as influências internas e externas. Na construção de uma sociedade economicamente próspera e justa é importante que as empresas possuam comportamento responsável e realizem estratégias em co-responsabilidade com o estado e a sociedade civil no combate à corrupção. Portanto, o estudo da gestão de compliance serve como pontapé inicial para a redução de riscos e práticas anti-éticas.
Tipo de documento	Artigo
Publicação em congresso	Não se aplica
<b>ARQUIVO</b>	
Link de acesso	<a href="#">PSP2_1_2018_Combate_a_corrupcao_em_empresas_privadas</a>
Tamanho	220 KB
Formato	Adobe PDF

Figura 5.4: Página Web legível para humanos.

Fonte: Adaptado de (Lóscio, Burle e Calegari 2017)<sup>4</sup>

<sup>4</sup>A programação CSS da página foi retirada de um exemplo de Boas Práticas para Dados na Web.

### 5.2.2.1 Aplicação dos requisitos

Levando em consideração as necessidades do PUMA, listou-se alguns metadados que poderão auxiliar na rastreabilidade e interoperabilidade dos dados de projetos. Os metadados estão na Tabela 5.2, assim como as descrições destes. A Tabela 5.3 mostra os termos e vocabulários associados aos metadados. Para a representação semântica, foram selecionados o vocabulário Dublin Core Metadata Initiative (DC), Friend of a Friend Vocabulary (FOAF) e Data Catalog Vocabulary (DCAT). As escolhas dos vocabulários se deram de acordo com o propósito de cada um deles. O Dublin Core é um vocabulário que tem por objetivo descrever objetos digitais, tais como textos, imagens e vídeos. Este vocabulário é amplamente utilizado para caracterizar publicações científicas, especialmente em sistemas de registro de produções intelectuais. O FOAF é dedicado a estabelecer ligações entre pessoas e informações na Web e o DCAT é um catálogo recomendado pelo W3C que visa aumentar a interoperabilidade entre vocabulários da Web e que facilita a pesquisa de um conjunto de dados.

Tabela 5.2: Definição e descrição dos metadados do Projeto PUMA.

<b>METADADO</b>	<b>DESCRIÇÃO</b>
Disciplinas vinculadas	Registro das disciplinas nas quais o trabalho foi desenvolvido
Título do Projeto	Denominação do trabalho
Área/Subárea	Área/Subárea da Engenharia de Produção segundo a Abepro <sup>5</sup>
Autor (es)	Estudantes desenvolvedores do projeto
Email dos autores e Professor (es) orientador (es)	Contatos dos responsáveis pelo documento
Data de apresentação	Data em que o trabalho foi apresentado na disciplina
Data de publicação	Data em que o trabalho foi registrado no PUMA
Professor (es) orientador (es)	Professor (es) responsável
Tipo de documento/resultado final	Resultado final do projeto (artigo, relatório, slides)
Palavras-chave	Palavras identificadoras do tema pesquisado
Resumo	Descrição breve do projeto realizado
Publicação em congresso	Registro de publicação em congresso
Link de acesso	Hiperlink para visualizar ou baixar o arquivo
Tamanho	Tamanho do arquivo em bytes
Formato	Formato do arquivo (PDF, CSS ou PPT por exemplo)

Fonte: Autoria própria.

---

<sup>5</sup><<http://portal.abepro.org.br/>>

Tabela 5.3: Termos e vocabulários dos metadados.

METADADO	TERMO	VOCABULÁRIO
Disciplinas vinculadas	dct:subject	Dublin core
Título do Projeto	dct:title	Dublin core
Área/Subárea	dct:subject	Dublin core
Autor (es)	dct:creator	Dublin core
Email dos autores e Professor (es) orientador (es)	foaf:mail	FOAF
Data de apresentação	dct:dateSubmitted	Dublin core
Data de publicação	dct:issued	Dublin core
Professor (es) orientador (es)	dc.contributor.advisor	Dublin core
Tipo de documento/resultado final	dcat:mediaType	DCAT
Palavras-chave	dcat:keyword	DCAT
Resumo	dct:description	Dublin core
Publicação em congresso	dc.publisher.meet	Dublin core
Link de acesso	dcat.downloadURL	DCAT
Tamanho	dcat.byteSize	DCAT
Formato	dcat.mediaType	DCAT

Fonte: Autoria própria.

Os metadados foram embutidos na Página Web através do modelo RDFa por ser de fácil implementação e leitura. De acordo com (Adida et al. 2012), o RFDa permite aos autores marcar dados, que são facilmente legíveis para pessoas e para máquinas, através da interpretação de navegadores e outros programas.

A Figura 5.5 mostra a programação HTML-RDFa da página apresentada na Figura 5.4.

```

<table class="rdfa-example">
  <tbody>
    <tr>
      <td>Disciplinas vinculadas</td>
      <td><span property="dct:subject">Projeto de Sistemas de Produção 2</span> <br /> <span property="dct:subject">Sistemas de Informação em Engenharia de Produção</span></td>
    </tr>
    <tr>
      <td>Área/Subárea</td>
      <td><span property="dct:subject">Engenharia de Produção, Sustentabilidade e Responsabilidade Social</span> <br> <span property="dct:subject">Ética e Transparência nas D</span></td>
    </tr>
    <tr>
      <td>Título</td>
      <td><span property="dct:title">Combate à corrupção em empresas privadas, questão de integridade e compliance: estudo de caso siemens</span></td>
    </tr>
    <tr>
      <td>Palavras-chave</td>
      <td><span property="dcat:keyword">Compliance</span> <br> <span property="dcat:keyword">Integridade</span> <br> <span property="dcat:keyword">Corrupção</span> <br> <span property="dcat:keyword">Integridade</span></td>
    </tr>
    <tr>
      <td>Autor(es)</td>
      <td><span property="dct:creator" typeof="foaf:Person"><span property="foaf:name">Pamella Oliveira de Souza</span> <small><a href="mailto:pamella.exemplo@unb.br" property="foaf:mail"></a></small></span> <span property="dct:creator" typeof="foaf:Person"><span property="foaf:name">Ada Maria Lopes Santos</span> <small><a href="mailto:ada.exemplo@unb.br" property="foaf:mail"></a></small></span> <span property="dct:creator" typeof="foaf:Person"><span property="foaf:name">Salém Miranda Alves</span> <small><a href="mailto:salem.exemplo@unb.br" property="foaf:mail"></a></small></span> <span property="dct:creator" typeof="foaf:Person"><span property="foaf:name">Yan Vieira de Jesus Costa</span> <small><a href="mailto:yan.costa@unb.br" property="foaf:mail"></a></small></span></td>
    </tr>
    <tr>
      <td>Orientador(es)</td>
      <td><span property="dc:contributor.advisor" typeof="foaf:Person"><span property="foaf:name">Edgard Costa Oliveira</span> <small><a href="mailto:edgard.exemplo@unb.br" property="foaf:mail"></a></small></span></td>
    </tr>
    <tr>
      <td>Data de apresentação</td>
      <td><span property="dct:dateSubmitted" content="2018-06-26" datatype="xsd:date">26 de junho de 2018</span></td>
    </tr>
    <tr>
      <td>Data de publicação</td>
      <td><span property="dct:issued" content="2021-04-08" datatype="xsd:date">08 de abril de 2021</span></td>
    </tr>
    <tr>
      <td>Resumo</td>
      <td><span property="dct:description">A corrupção é definida como abuso público ou privado de poder para obter benefício próprio, para amigos ou familiares, está presente em t</span></td>
    </tr>
    <tr>
      <td>Tipo de documento/formato</td>
      <td><span property="dcat:mediaType">Artigo</span> <br> <span property="dcat:mediaType">PDF</span></td>
    </tr>
  </tbody>
</table>

```

Figura 5.5: Programação RDFa para registros de Recursos Educacionais Abertos.

Fonte: Autoria própria.

Um campo importante adicionado foi um convite para que os usuários dos dados deixem seus *feedbacks* (Figura 5.6). Esses *feedbacks* podem ser registrados em um formulário de contato simples no próprio Website do PUMA. Embora não seja possível usufruir das contribuições dos usuários durante a criação da plataforma, é importante considerar que a geração de dados será constante dentro do PUMA, posto que a plataforma irá passar por melhorias contínuas, visando a qualidade do sistema.

A licença utilizada foi a Creative Commons, como mostra a Figura 5.6. Essa licença foi escolhida por ser amplamente utilizada globalmente. Com ela, pode-se copiar, alterar e distribuir o conteúdo, portanto que seja dada a devida atribuição, o conteúdo não seja utilizado para fins lucrativos e que o usuário compartilhe com a mesma licença (Creative Commons 2021).

**ARQUIVO**

Link de acesso	<a href="#">PSP2_1_2018_Combate_a_corrupcao_em_empresas_privadas</a>
Tamanho	220 KB
Formato	Adobe PDF

**FEEDBACK**

Dúvidas ou sugestões? Por favor deixe seu feedback no nosso [Formulário de Contato](#).


  
Este obra está licenciado com uma Licença [Creative Commons 4.0 Internacional](#).

Figura 5.6: Formato de registro de feedback e licença do item.

Fonte: Autoria própria.

O documento web pode ser acessado através da URI <<http://pamellaprodu.github.io/>>.

Os vocabulários, termos, representação sintática dos dados e licença escolhidos visam apoiar e possibilitar a interoperabilidade dos datasets da plataforma e promover a qualidade dos dados gerados ao aplicar padrões internacionalmente conceituados, de acordo com os Princípios para Dados Abertos, as Melhores Práticas para Publicação de Dados Conectados e as Boas Práticas para Dados na Web.

## 6 CONSIDERAÇÕES FINAIS, LIMITAÇÕES E FUTURAS LINHAS DE PESQUISA

O presente trabalho teve por objetivo o levantamento de requisitos que possibilitem a geração de Recursos Educacionais Abertos Conectados, ou seja, recursos educacionais que contemplassem os princípios dos Dados Abertos, Dados Conectados e que seguissem Boas Práticas para a publicação de dados na Web, com o propósito de torná-los interoperáveis e favorecer a qualidade da informação.

No primeiro momento do projeto, buscou-se compreender a problemática através de pesquisas iniciais e entrevistas com representantes de projetos educacionais, em especial o PUMA. Após uma pesquisa bibliográfica orientada por uma revisão sistemática da literatura usando o método TEMAC, detalhou-se as Boas Práticas para Dados na Web, uma recomendação W3. Também aprofundou-se em vocabulários internacionalmente conceituados e em modelos de representação sintática para Dados Conectados.

A partir desse arcabouço teórico, foi possível elencar alguns requisitos que corroboram para a criação de Recursos Educacionais Abertos e Conectados com outros dados na Web. Para a elaboração desses requisitos, foram considerados os Princípios para Dados Abertos, Melhores Práticas para Publicação de Dados Conectados e Boas Práticas para Dados na Web. Esses requisitos foram: fornecimento de metadados básicos e descritivos, fornecimento de diversos formatos de dados, priorizando os processáveis por máquinas e não proprietários, uso de vocabulários padronizados, fornecimento de licença, uso de bons URIs e coleta de *feedbacks*. Ressalta-se que o *feedback* é primordial para melhoria contínua dos dados gerados em projetos educacionais.

Por fim, criou-se um modelo simples baseado nas necessidades do projeto PUMA para a criação de seu banco de dados. Definiu-se metadados, vocabulários, representação sintática no modelo RDF, tipo de licença e uso de um formulário para aquisição de *feedbacks* dos usuários na plataforma. O documento está online por meio de um URI permanente.

Através das execução das etapas supracitadas, foi possível cumprir com os objetivos elencados para o projeto. O modelo desenvolvido tem potencial de apoiar não apenas o projeto PUMA, mas qualquer projeto educacional que tenha interesse em preparar seus dados, vislumbra a integração destes com a Web.

### 6.1 LIMITAÇÕES DA PESQUISA

A pesquisa realizada teve como fator limite o acesso aos resultados gerados nas disciplinas de Projetos de Sistemas de Produção (PSPs), o que impactou na diminuição do escopo da proposta de representação semântica detalhada durante a subseção 5.2.2. A priori pensou-se em criar um banco de dados com vários projetos. A falta de dados prejudica os estudos iniciais de ferramentas e plataformas para a extração de valor dos projetos desenvolvidos dentro do curso de Engenharia de Produção.

## 6.2 FUTURAS LINHAS DE PESQUISA

Há inúmeras possibilidades de desenvolvimento de futuros trabalhos, em especial da área na Computação Aplicada. As Boas Práticas para Dados na Web recomendam, por exemplo, promover o acesso aos dados por meio de download em massa e fornecimento de subconjuntos de dados. Ademais, é essencial para a interoperabilidade que haja o desenvolvimento de APIs, possibilitando que usuários usufruam dos recursos educacionais com mais facilidade.

Espera-se que este trabalho seja apenas o primeiro, dentro do Curso de Engenharia de Produção, de uma série de estudos que venham aplicar recomendações para a integração de dados educacionais na Web, posto que a Educação é a ferramenta mais poderosa que existe para a equidade social.



# REFERÊNCIAS BIBLIOGRÁFICAS

A3M 2018 A3M. *Plataforma Unificada de Metodologia Ativa (PUMA)*. 2018. Disponível em: <<https://www.a3m.cead.unb.br/projetos/plataforma-unificada-de-metodologia-ativa-puma>>.

A3M 2021 A3M. *Aprendizagem para o 3º Milênio*. 2021. Disponível em: <<https://www.a3m.cead.unb.br/>>.

Abramo e D'Angelo 2011 ABRAMO, G.; D'ANGELO, C. A. Evaluating research: from informed peer review to bibliometrics. *Scientometrics*, Akadémiai Kiadó, co-published with Springer Science+ Business Media BV ..., v. 87, n. 3, p. 499–514, 2011. Disponível em: <<https://doi.org/10.1007/s11192-011-0352-7>>.

Adida et al. 2015 ADIDA, B.; BIRBECK, M.; MCCARRON, S.; HERMAN, I. *RDFa Core 1.1 - Third Edition - Syntax and processing rules for embedding RDF through attributes*. 2015. Disponível em: <<https://www.w3.org/TR/rdfa-core/>>.

Adida et al. 2012 ADIDA, B.; HERMAN, I.; SPORNY, M.; BIRBECK, M. *Manual de RDFa 1.1*. 2012. Disponível em: <<https://skos.um.es/TR/rdfa-primer/>>.

Aguiar e Medeiros 2009 AGUIAR, V. R. L.; MEDEIROS, C. M. Entrevistas na pesquisa social: o relato de um grupo de foco nas licenciaturas. *Congr. Nac. Educação, EDUCERE*, v. 9, p. 10–710, 2009.

Albertoni et al. 2020 ALBERTONI, R.; BROWNING, D.; COX, S.; BELTRAN, A.; PEREGO, A.; WINSTANLEY, P. *Vocabulário do Catálogo de Dados (DCAT) - Versão 3*. 2020. Disponível em: <<https://www.w3.org/TR/vocab-dcat-3/>>.

Albrecht et al. 2017 ALBRECHT, C.-M.; BACKHAUS, C.; GURZKI, H.; WOISETSCHLÄGER, D. M. Value creation for luxury brands through brand extensions. In: *Luxusmarkenmanagement*. Springer, 2017. p. 261–283. Disponível em: <[https://doi.org/10.1007/978-3-658-09072-2\\_13](https://doi.org/10.1007/978-3-658-09072-2_13)>.

Alcantara et al. 2015 ALCANTARA, W.; BANDEIRA, J.; BARBOSA, A.; LIMA, A.; ÁVILA, T.; BITTERCOURT, I.; ISOTANI, S. Desafios no uso de dados abertos conectados na educação brasileira. In: SBC. *Anais do IV Workshop de Desafios da Computação Aplicada a Educação*. [S.l.], 2015. p. 11–20.

Amiel, Gonsales e Sebriam 2018 AMIEL, T.; GONSALES, P.; SEBRIAM, D. Recursos educacionais abertos no Brasil: 10 anos de ativismo. *EmRede-Revista de Educação a Distância*, v. 5, n. 2, p. 246–258, 2018.

Arenas, García e Espasandín 2001 ARENAS, J.; GARCÍA, R.; ESPASANDIN, F. Aproximación empírica sobre el análisis de la literatura de alianzas estratégicas. In: *Proceedings of X International Conference of AEDEM*. [S.l.: s.n.], 2001.

Armentano, Godoy e Amandi 2012 ARMENTANO, M. G.; GODOY, D.; AMANDI, A. Topology-based recommendation of users in micro-blogging communities. *Journal of Computer Science and Technology*, Springer, v. 27, n. 3, p. 624–634, 2012.

Atenas et al. 2015 ATENAS, J. et al. Open data as open educational resources: Towards transversal skills and global citizenship. *Open praxis*, International Council for Open and Distance Education, v. 7, n. 4, p. 377–389, 2015.

- Bandeira et al. 2014 BANDEIRA, J.; ALCANTARA, W.; BARBOSA, A.; ÁVILA, T.; BITTENCOURT, I.; ISOTANI, S. Dados abertos conectados na educação brasileira. *Jornada de Atualização em Tecnologia da Informação. Anais do III Simpósio Brasileiro de Tecnologia da Informação-SBTI*, 2014.
- Bandeira et al. 2015 BANDEIRA, J.; ÁVILA, T.; ALCANTARA, W.; SOBRINHO, A.; BITTENCOURT, I. I.; ISOTANI, S. Dados abertos conectados para a educação. *Jornada de Atualização em Informática na Educação*, v. 4, n. 1, p. 47–69, 2015.
- Batista e Lóscio 2013 BATISTA, M. G. R.; LÓSCIO, B. F. Opensbbd: Usando linked data para publicação de dados abertos sobre o sbbd. In: *SBBD (Short Papers)*. [S.l.: s.n.], 2013. p. 10–1.
- Beckett e Berners-Lee 2011 BECKETT, D.; BERNERS-LEE, T. *Turtle - Terse RDF Triple Language*. 2011. Disponível em: <<https://www.w3.org/TR/turtle/>>.
- Berners-Lee 2006 Berners-Lee. *W3C - Design Issues*. 2006. Disponível em: <<https://www.w3.org/DesignIssues/>>.
- Berners-Lee 2006 BERNERS-LEE, T. *Design Issues - Linked Data*. 2006. Disponível em: <<https://www.w3.org/DesignIssues/LinkedData>>.
- Berners-Lee 2010 BERNERS-LEE, T. *As 5 estrelas dos dados abertos*. 2010. Disponível em: <<https://5stardata.info/pt-BR/>>.
- Berners-Lee, Hendler e Lassila 2001 BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific american*, JSTOR, v. 284, n. 5, p. 34–43, 2001.
- Bittencourt et al. 2008 BITTENCOURT, I. I.; ISOTANI, S.; COSTA, E.; MIZOGUCHI, R. Research directions on semantic web and education. *Interdisciplinary Studies in Computer Science*, Citeseer, v. 19, n. 1, p. 60–67, 2008.
- Bizer et al. 2007 BIZER, C.; CYGANIAK, R.; HEATH, T. et al. How to publish linked data on the web. 2007.
- Bizer, Heath e Berners-Lee 2011 BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data: The story so far. In: *Semantic services, interoperability and web applications: emerging concepts*. [S.l.]: IGI Global, 2011. p. 205–227.
- Brickley e Mille 2014 BRICKLEY, D.; MILLE, L. *FOAF Vocabulary Specification 0.99*. 2014. Disponível em: <<http://xmlns.com/foaf/0.1/>>.
- Calazans, MASSON e MARIANO 2015 CALAZANS, A. T. S.; MASSON, E. T. S.; MARIANO, A. M. Uma revisão sistemática da bibliografia sobre inovação bancária utilizando o enfoque meta-analítico. *Revista ESPACIOS| Vol. 36 (Nº 15) Año 2015*, 2015.
- Carothers e Seaborne 2014 CAROTHERS, G.; SEABORNE, A. *RDF 1.1 TriG*. 2014. Disponível em: <<https://www.w3.org/TR/trig/>>.
- Chang e Blei 2009 CHANG, J.; BLEI, D. Relational topic models for document networks. In: *Artificial Intelligence and Statistics*. [S.l.: s.n.], 2009. p. 81–88.
- Chicaiza et al. 2014 CHICAIZA, J.; PIEDRA, N.; LOPEZ-VARGAS, J.; TOVAR-CARO, E. Domain categorization of open educational resources based on linked data. In: SPRINGER. *International Conference on Knowledge Engineering and the Semantic Web*. [S.l.], 2014. p. 15–28.
- Chignard 2013 CHIGNARD, S. A brief history of open data. *Paris Tech Review*, v. 29, 2013.

Ciancarini, Poggi e Russo 2016 CIANCARINI, P.; POGGI, F.; RUSSO, D. Big data quality: a roadmap for open data. In: IEEE. *2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService)*. [S.l.], 2016. p. 210–215.

Coughlan 2020 COUGHLAN, T. The use of open data as a material for learning. *Educational Technology Research and Development*, Springer, v. 68, n. 1, p. 383–411, 2020.

Creative Commons 2021 Creative Commons. *Share your work*. 2021. Disponível em: <<https://creativecommons.org/share-your-work/>>.

Cruz e Correa 2004 CRUZ, R. G.; CORREA, P. R. El meta análisis como instrumento de investigación en la determinación y análisis del objeto de estudio. *XVI Encuentro de Profesores Universitarios de Marketing (2004)*, p 1-16, 2004. Disponível em: <<https://hdl.handle.net/11441/71853>>.

Curry et al. 2013 CURRY, E.; O'DONNELL, J.; CORRY, E.; HASAN, S.; KEANE, M.; O'RIAIN, S. Linking building data in the cloud: Integrating cross-domain building data using linked data. *Advanced Engineering Informatics*, Elsevier, v. 27, n. 2, p. 206–219, 2013.

Czerkowski 2014 CZERKAWSKI, B. Ö. The semantic web in teacher education. *Turkish Online Journal of Educational Technology-TOJET*, ERIC, v. 13, n. 4, p. 144–147, 2014.

Daga et al. 2016 DAGA, E.; D'AQUIN, M.; ADAMOU, A.; BROWN, S. The open university linked data–data. open. ac. uk. *Semantic Web*, IOS Press, v. 7, n. 2, p. 183–191, 2016.

Delfino, Neto e Sousa 2019 DELFINO, S. S.; NETO, J. A. S. de P.; SOUSA, M. R. F. de. Desafios da sociedade da informação na recuperação e uso de informações em ambientes digitais. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, v. 17, p. e019036–e019036, 2019. Disponível em: <<https://doi.org/10.20396/rdbci.v17i0.8655973>>.

Deshpande e Huang 2011 DESHPANDE, A. A.; HUANG, S. H. Simulation games in engineering education: A state-of-the-art review. *Computer applications in engineering education*, Wiley Online Library, v. 19, n. 3, p. 399–410, 2011.

Dietze et al. 2013 DIETZE, S.; SANCHEZ-ALONSO, S.; EBNER, H.; YU, H. Q.; GIORDANO, D.; MARENZI, I.; NUNES, B. P. Interlinking educational resources and the web of data. *Program*, Emerald Group Publishing Limited, 2013. Disponível em: <<https://doi.org/10.1108/00330331211296312>>.

Dublin Core Metadata Initiative 2021 Dublin Core Metadata Initiative. *DCMI Metadata Terms*. 2021. Disponível em: <<https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>>.

D'Antoni 2009 D'ANTONI, S. *Open educational resources: Reviewing initiatives and issues*. [S.l.]: Taylor & Francis, 2009.

Fagundes, Macedo e Freund 2018 FAGUNDES, P. B.; MACEDO, D. D. J. de; FREUND, G. P. A produção científica sobre qualidade de dados em big data: um estudo na base de dados web of science. *RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação*, v. 16, n. 1, p. 194–210, 2018. Disponível em: <<https://doi.org/10.20396/rdbci.v16i1.8650412>>.

Farias et al. 2015 FARIAS, P. A. M. d.; MARTIN, A. L. d. A. R.; CRISTO, C. S. et al. Aprendizagem ativa na educação em saúde: percurso histórico e aplicações. *Revista brasileira de educação médica*, SciELO Brasil, v. 39, n. 1, p. 143–150, 2015.

Ferreira e Segurado 2019 FERREIRA, J. E.; SEGURADO, A. C. 2. interoperabilidade de dados, desempenho acadêmico e impacto social: Usp no horizonte 2022. *REPENSAR A UNIVERSIDADE II*, p. 43, 2019.

FGV DAPP 2020 FGV DAPP. *Modelo de Referência de Abertura de Dados Abertos*. 2020. Disponível em: <<https://wiki.dados.gov.br/Politica-de-Dados-Abertos.ashx>>.

Floridi 2010 FLORIDI, L. *Information: A very short introduction*. [S.l.]: OUP Oxford, 2010.

Fürber e Hepp 2010 FÜRBER, C.; HEPP, M. Using sparql and spin for data quality management on the semantic web. In: SPRINGER. *International Conference on Business Information Systems*. [S.l.], 2010. p. 35–46.

Geraci et al. 1991 GERACI, A.; KATKI, F.; MCMONEGAL, L.; MEYER, B.; LANE, J.; WILSON, P.; RADATZ, J.; YEE, M.; PORTEOUS, H.; SPRINGSTEEL, F. *IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries*. [S.l.]: IEEE Press, 1991.

Gil 2017 GIL, A. C. *Como elaborar projetos de pesquisa*. [S.l.]: Atlas São Paulo, 2017. v. 6.

Gil e Miles 2013 GIL, Y.; MILES, S. *PROV Model Primer*. 2013. Disponível em: <<https://www.w3.org/TR/2013/NOTE-prov-primer-20130430/>>.

Governo Digital 2019 Governo Digital. *Infraestrutura Nacional de Dados Abertos*. 2019. Disponível em: <<https://www.gov.br/>>.

Guterres 2020 GUTERRES, A. *ONU: mundo deve 'redesenhar' a educação em meio à pandemia*. 2020. Disponível em: <<https://news.un.org/>>.

Herman et al. 2015 HERMAN, I.; ADIDA, B.; SPORNY, M.; BIRBECK, M. *RDFa 1.1 Primer - Third Edition*. 2015. Disponível em: <<https://www.w3.org/TR/rdfa-primer/>>.

Hyland, Ateazing e Villazón-Terrazas 2014 HYLAND, B.; ATEMEZING, G.; VILLAZÓN-TERRAZAS, B. *Best Practices for Publishing Linked Data*. 2014. Disponível em: <<https://www.w3.org/TR/ld-bp/>>.

IDC Infobrief 2020 IDC Infobrief. *Using Data Intelligently: Unification and Pipelining Patterns in the Digital Economy*. 2020. Disponível em: <<https://www.tibco.com/resources/analyst-report/using-data-intelligently>>.

Isotani e Bittencourt 2015 ISOTANI, S.; BITTENCOURT, I. *Dados Abertos Conectados: em Busca da Web do Conhecimento*. [s.n.], 2015. ISBN 978-85-7522-449-6. Disponível em: <<https://doi.org/10.13140/RG.2.1.4355.6329>>.

Janssen, Charalabidis e Zuiderwijk 2012 JANSSEN, M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. Benefits, adoption barriers and myths of open data and open government. *Information systems management*, Taylor & Francis, v. 29, n. 4, p. 258–268, 2012.

Jevsikova et al. 2017 JEVSIKOVA, T. et al. Application of resource description framework to personalise learning: Systematic review and methodology. *Informatics in Education*, Vilnius University Institute of Mathematics and Informatics, Lithuanian ... , v. 16, n. 1, p. 61–82, 2017. Disponível em: <<https://doi.org/10.1016/j.aei.2012.10.003>>.

Júnior et al. 2019 JÚNIOR, E. S.; MONTEIRO, S. B. S.; LIMA, A. C. F.; MARIANO, A. M.; SILVA, V. B. N. da. Unified platform of active methodology (puma): Development steps of the module divulgation and request projects. v. 9, p. 358–365, 2019.

Laufer 2015 LAUFER, C. Guia de web semântica. *Guia de web semântica*., 2015. Disponível em: <<https://ceweb.br/guias/web-semantica/>>.

- Leung, Sun e Bai 2017 LEUNG, X. Y.; SUN, J.; BAI, B. Bibliometrics of social media research: A co-citation and co-word analysis. *International Journal of Hospitality Management*, Elsevier, v. 66, p. 35–45, 2017.
- Love et al. 2016 LOVE, M.; BOISVERT, C.; URUCHURTU, E.; IBBOTSON, I. Nifty with data: can a business intelligence analysis sourced from open data form a nifty assignment? In: *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education*. [S.l.: s.n.], 2016. p. 344–349.
- Lóscio, Burle e Calegari 2017 LÓSCIO, B.; BURLE, C.; CALEGARI, N. *W3C - Data on the Web Best Practices*. 2017. Disponível em: <<https://www.w3.org/Translations/DWBP-pt-BR/>>.
- Mariano, Cruz e Gaitán 2011 MARIANO, A. M.; CRUZ, R. G.; GAITÁN, J. A. Meta análises como instrumento de pesquisa: Uma revisão sistemática da bibliografia aplicada ao estudo das alianças estratégicas internacionais. 2011. Disponível em: <<https://doi.org/10.13140/RG.2.1.1150.1522>>.
- Mariano e Rocha 2017 MARIANO, A. M.; ROCHA, M. S. Revisão da literatura: apresentação de uma abordagem integradora. In: *AEDEM International Conference*. [s.n.], 2017. v. 18. Disponível em: <<https://www.pesquisatemac.com/>>.
- Marquesone 2016 MARQUESONE, R. *Big Data: Técnicas e tecnologias para extração de valor dos dados*. [S.l.]: Editora Casa do Código, 2016.
- Martins, Mello e Turrioni 2013 MARTINS, R. A.; MELLO, J. B. P.; TURRIONI, C. H. *Guia para elaboração de monografia e TCC em engenharia de produção*. [S.l.]: Editora Atlas SA, 2013.
- May 2004 MAY, T. *Pesquisa social: questões, métodos e processos*. [S.l.]: Artmed, 2004.
- McKinsey Global Institute analysis 2013 McKinsey Global Institute analysis. *Open data: Unlocking innovation and performance with liquid information*. 2013. Disponível em: <<https://www.mckinsey.com/>>.
- Monteiro et al. 2017 MONTEIRO, S. B. S.; REIS, A. C. B.; SILVA, J. M. d.; SOUZA, J. C. F. A project-based learning curricular approach in a production engineering program. *Production*, SciELO Brasil, v. 27, n. SPE, 2017.
- Nahhas et al. 2018 NAHHAS, S.; BAMASAG, O.; KHEMAKHEM, M.; BAJNAID, N. Linked data approach to mutually enrich traditional education resources with global open education. In: *IEEE. 2018 1st International Conference on Computer Applications & Information Security (ICCAIS)*. [S.l.], 2018. p. 1–8.
- OECD 2007 OECD. *Giving knowledge for free: The emergence of open educational resources*. [S.l.]: Organisation for Economic Co-operation and Development, 2007.
- ONU Brasil 2012 ONU Brasil. *[Objetivos de Desenvolvimento Sustentável]*. 2012. Disponível em: <<https://brasil.un.org/pt-br/sdgs>>.
- Open Data Charter 2015 Open Data Charter. *Principles*. 2015. Disponível em: <<https://opendatacharter.net/principles/>>.
- Open Data Charter 2020 Open Data Charter. *Open Data Charter - Our history*. 2020. Disponível em: <<https://opendatacharter.net/our-history/>>.
- Open Government Partnership 2021 Open Government Partnership. *Members*. 2021. Disponível em: <<https://www.opengovpartnership.org/our-members/>>.

Open Knowledge Brasil 2017 Open Knowledge Brasil. *Dados Conectados*. 2017. Disponível em: <<https://www.ok.org.br/noticia/dados-conectados/>>.

Open Knowledge Brasil 2017 Open Knowledge Brasil. *Melhores Práticas para Publicação de Dados Conectados*. 2017. Disponível em: <<https://www.ok.org.br/noticia/melhores-praticas-para-publicacao-de-dados-conectados/#>>.

Penteado e Isotani 2020 PENTEADO, B.; ISOTANI, S. Modelo de processo com controle de qualidade para a produção de dados abertos conectados governamentais de alta qualidade. *IEEE Latin America Transactions*, v. 100, n. 1e, 2020. Disponível em: <<http://www.repositorio.ufal.br/handle/riufal/2209>>.

Penteado 2020 PENTEADO, B. E. *Modelo de infraestrutura para publicação de dados abertos governamentais conectados de qualidade*. Tese (Doutorado) — Universidade de São Paulo, 2020. Disponível em: <<https://doi.org/10.11606/T.55.2020.tde-14092020-175138>>.

Penteado, Bittencourt e Isotani 2019 PENTEADO, B. E.; BITTENCOURT, I. I.; ISOTANI, S. Análise exploratória sobre a abertura de dados educacionais no Brasil: como torná-los prontos para o ecossistema da web? *Revista Brasileira de Informática na Educação*, v. 27, n. 01, p. 175, 2019.

Pereira et al. 2018 PEREIRA, A. S.; SHITSUKA, D. M.; PARREIRA, F. J.; SHITSUKA, R. *Metodologia da pesquisa científica*. Brasil, 2018.

Piedra et al. 2014 PIEDRA, N.; CHICAIZA, J.; LÓPEZ, J.; CARO, E. T. Supporting openness of moocs contents through of an oer and ocr framework based on linked data technologies. In: IEEE. *2014 IEEE Global Engineering Education Conference (EDUCON)*. [S.l.], 2014. p. 1112–1117.

Piedra et al. 2015 PIEDRA, N.; CHICAIZA, J.; LÓPEZ-VARGAS, J.; CARO, E. T. Seeking open educational resources to compose massive open online courses in engineering education an approach based on linked open data. *J. UCS*, v. 21, n. 5, p. 679–711, 2015.

Piedra et al. 2016 PIEDRA, N.; CHICAIZA, J.; LOPEZ-VARGAS, J.; CARO, E. T. Guidelines to producing structured interoperable data from open access repositories. In: IEEE. *2016 IEEE Frontiers in Education Conference (FIE)*. [S.l.], 2016. p. 1–9.

Piedra, López e Chicaiza PIEDRA, N. T.; LÓPEZ, E.; CHICAIZA, J. J.(2014). *Consuming and producing linked Open Data: The case of OpenCourseWare*. [S.l.]: Emerald EarlyCite.

Portal Brasileiro de Dados Abertos 2015 Portal Brasileiro de Dados Abertos. *O que são dados abertos?* 2015. Disponível em: <<https://dados.gov.br/pagina/dados-abertos>>.

Portal Brasileiro de Dados Abertos 2019 Portal Brasileiro de Dados Abertos. *Cartilha Técnica para Publicação de Dados Abertos no Brasil*. 2019. Disponível em: <<https://dados.gov.br/pagina/cartilha-publicacao-dados-abertos>>.

Portal Brasileiro de Dados Abertos 2019 Portal Brasileiro de Dados Abertos. *Modelo de Maturidade de publicação de dados*. 2019. Disponível em: <<https://wiki.dados.gov.br/>>.

Portal Brasileiro de Dados Abertos 2020 Portal Brasileiro de Dados Abertos. *INFRAESTRUTURA NACIONAL DE DADOS ABERTOS (INDA)*. 2020. Disponível em: <<https://wiki.dados.gov.br/>>.

Portal da Transparência 2018 Portal da Transparência. *Portal da Transparência - O que é e como funciona*. 2018. Disponível em: <<http://www.portaltransparencia.gov.br/>>.

PORTAL DE PERIÓDICOS DA CAPES/MEC 2020 PORTAL DE PERIÓDICOS DA CAPES/MEC. *Web of Science - Coleção Principal*. 2020. Disponível em: <<http://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/>>.

- Prud'hommeaux e Carothers 2014 PRUD'HOMMEAUX, E. P.; CAROTHERS, G. *RDF 1.1 Turtle - Terse RDF Triple Language*. 2014. Disponível em: <<https://www.w3.org/TR/turtle/>>.
- Rautenberg et al. 2017 RAUTENBERG, S.; MOTYL, S. K.; BURDA, A. C.; SILVERIO, A.; MOURA, F. M. de. Linked open data and knowledge management: scientometric study cases in a brazilian university. *PERSPECTIVAS EM CIENCIA DA INFORMACAO*, UNIV FEDERAL MINAS GERAIS, ESCOLA BIBLIOTECONOMIA CENTRO, CAIXA POSTAL 1606 . . . , v. 22, n. 3, p. 116–+, 2017.
- Silva e Menezes 2005 SILVA, E. L. d.; MENEZES, E. M. Metodologia da pesquisa e elaboração de dissertação. rev. atual. *Florianópolis: UFSC*, 2005.
- Silva e Balthazar 2010 SILVA, J. M. da; BALTHAZAR, J. C. *Projeto Pedagógico do Curso de Engenharia de Produção*. 2010. Disponível em: <<http://www.epr.unb.br/>>.
- Sporny, et al. 2020 SPORNY, M.; et al. *JSON-LD 1.1 - A JSON-based Serialization for Linked Data*. 2020. Disponível em: <<https://www.w3.org/TR/json-ld11/>>.
- Tomoyose 2021 TOMOYOSE, K. O data catalog vocabulary (dcat) para a publicação de dados de pesquisa nos princípios linked data. Universidade Federal de São Carlos, 2021. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/14116>>.
- Tovar e Piedra 2014 TOVAR, E.; PIEDRA, N. Guest editorial: open educational resources in engineering education: various perspectives opening the education of engineers. *IEEE Transactions on Education*, IEEE, v. 57, n. 4, p. 213–219, 2014.
- UNESCO 2012 UNESCO. *Paris OER Declaration, 2012 World Open Educational Resources (OER) Congress*. [S.l.]: UNESCO, Paris, 2012.
- Valente e Fujino 2016 VALENTE, N. T. Z.; FUJINO, A. Atributos e dimensões de qualidade da informação nas ciências contábeis e na ciência da informação: um estudo comparativo. *Perspectivas em Ciência da Informação*, SciELO Brasil, v. 21, n. 2, p. 141–167, 2016.
- Vogel e Güttel 2013 VOGEL, R.; GÜTTEL, W. H. The dynamic capability view in strategic management: A bibliometric review. *International Journal of Management Reviews*, Wiley Online Library, v. 15, n. 4, p. 426–446, 2013.
- W3C 2012 W3C. *Introduction to SKOS*. 2012. Disponível em: <<https://www.w3.org/2004/02/skos/intro>>.
- W3C 2014 W3C. *W3C Recommendations - RDF Working Group*. 2014. Disponível em: <[https://www.w3.org/2011/rdf-wg/wiki/Main\\_Page](https://www.w3.org/2011/rdf-wg/wiki/Main_Page)>.
- W3C 2015 W3C. *Ontologies*. 2015. Disponível em: <<https://www.w3.org/standards/semanticweb/ontology>>.
- W3C 2015 W3C. *Semantic Web*. 2015. Disponível em: <<https://www.w3.org/standards/semanticweb/>>.
- W3C 2015 W3C. *W3C - Semantic Web*. 2015. Disponível em: <<https://www.w3.org/standards/semanticweb/>>.
- W3C 2016 W3C. *Wiki - Linked Data*. 2016. Disponível em: <<https://www.w3.org/wiki/LinkedData>>.
- W3C 2020 W3C. *RDFa Core Initial Context*. 2020. Disponível em: <<https://www.w3.org/2011/rdfa-context/rdfa-1.1>>.
- W3Schools 2021 W3Schools. *XML RDF*. 2021. Disponível em: <[https://www.w3schools.com/xml/xml\\_rdf.asp](https://www.w3schools.com/xml/xml_rdf.asp)>.

Wang e Strong 1996 WANG, R. Y.; STRONG, D. M. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, Taylor & Francis, v. 12, n. 4, p. 5–33, 1996.

Wood et al. 2014 WOOD, D.; ZAIDMAN, M.; RUTH, L.; HAUSENBLAS, M. *Linked Data*. [S.l.]: Manning publications co., 2014.

World Bank 2019 World Bank. *Open Government Data Toolkit*. 2019. Disponível em: <<http://opendatatoolkit.worldbank.org/>>.

World Wide Web Foundation 2018 World Wide Web Foundation. *Open Data Barometer – Leaders Edition*. 2018. Disponível em: <<https://opendatabarometer.org/leadersedition/report/>>.

Worthy 2015 WORTHY, B. The impact of open data in the uk: Complex, unpredictable, and political. *Public Administration*, Wiley Online Library, v. 93, n. 3, p. 788–805, 2015.

Zaveri et al. 2016 ZAVERI, A.; RULA, A.; MAURINO, A.; PIETROBON, R.; LEHMANN, J.; AUER, S. Quality assessment for linked data: A survey. *Semantic Web*, IOS Press, v. 7, n. 1, p. 63–93, 2016.

Zindel et al. 2012 ZINDEL, M. L.; SILVA, J. Mello da; SOUZA, J.; MONTEIRO, S.; OLIVEIRA, E. A new approach in engineering education: the design-centric curriculum at the university of brasília-brazil. *International Journal of Basic & Applied Sciences*, Citeseer, v. 12, n. 5, p. 97–102, 2012. Disponível em: <[http://www.ijens.org/Vol\\_12\\_I\\_05/127105-8585-IJBAS-IJENS.pdf](http://www.ijens.org/Vol_12_I_05/127105-8585-IJBAS-IJENS.pdf)>.



## APÊNDICES

# I. PUBLICAÇÕES E CITAÇÕES SOBRE O TEMA RECURSOS EDUCACIONAIS ABERTOS

As Figuras I.1 e I.2 apresentam os dados de publicações e citações sobre o tema "*Open Educational Resources*". Foi considerado o período 2005-2020.

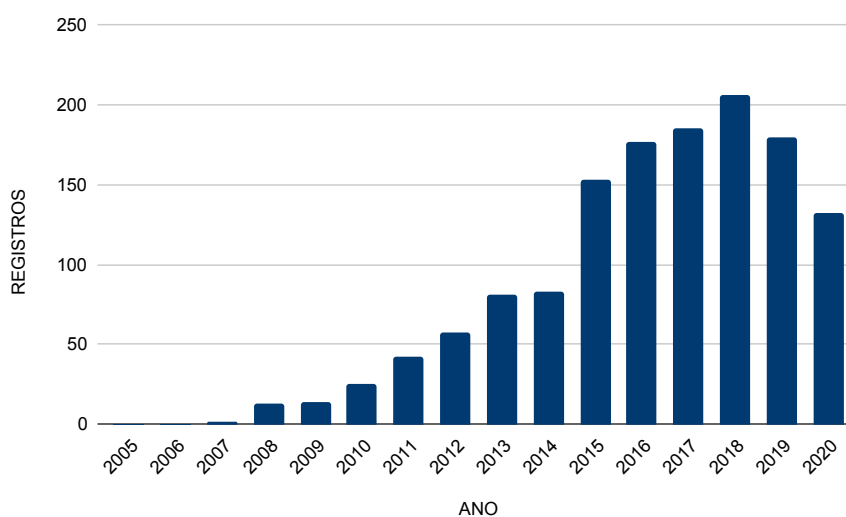


Figura I.1: Número de publicações ano a ano para Recursos Educacionais Abertos.

Fonte: Autoria própria com base em dados do *ISI Web of Science*.

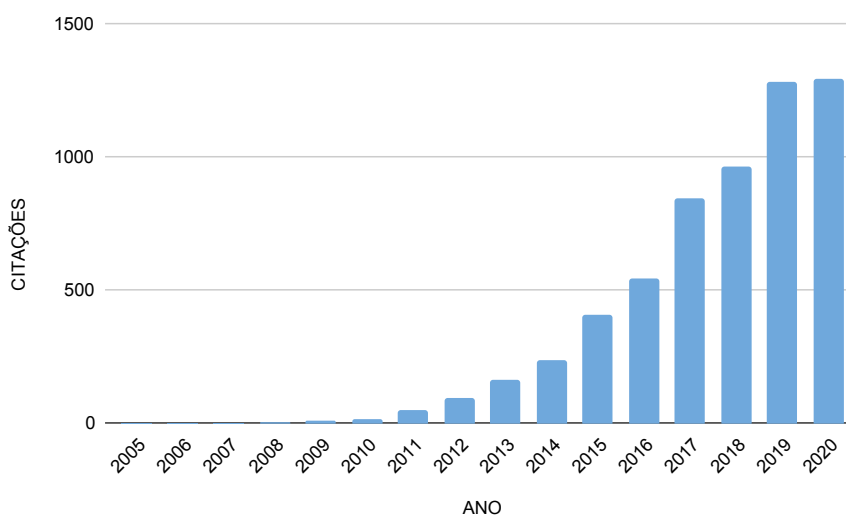


Figura I.2: Número de citações ano a ano para Recursos Educacionais Abertos.

Fonte: Autoria própria com base em dados do *ISI Web of Science*.