



Pós-Graduação em Ciência da Computação

Rayelle Ingrid Vera Cruz Silva Muniz

DWMM: Um Modelo de Maturidade para Dados na Web



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
<http://cin.ufpe.br/~posgraduacao>

Recife
2019

Rayelle Ingrid Vera Cruz Silva Muniz

DWMM: Um Modelo de Maturidade para Dados na Web

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Banco de Dados

Orientador: Profa. Dra. Bernadette Farias Lóscio

Recife
2019

Catálogo na fonte
Bibliotecária Arabelly Ascoli CRB4-2068

M966d Muniz, Rayelle Ingrid Vera Cruz Silva
 DWMM: um modelo de maturidade para dados na web /
 Rayelle Ingrid Vera Cruz Silva Muniz . – 2019.
 118 f.: fig., tab.

 Orientadora: Bernadette Farias Lóscio
 Dissertação (Mestrado) – Universidade Federal de
 Pernambuco. Cln. Ciência da Computação. Recife, 2019.
 Inclui referências e apêndice.

 1. Dados na web. 2. Publicação de dados. 3. Modelo de
 maturidade. 4. Ciclo de vida. I. Lóscio, Bernadette Farias
 (orientadora). II. Título.

025.04

CDD (22. ed.)

UFPE-CCEN 2020-28

Rayelle Ingrid Vera Cruz Silva Muniz

“DWMM: Um Modelo de Maturidade para Dados na Web”

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

Aprovado em: 26 de agosto de 2019.

BANCA EXAMINADORA

Prof. Dr. Kiev dos Santos Gama
Centro de Informática / UFPE

Prof. Dr. Carlos Denner dos Santos Júnior
Departamento de Administração / UnB

Profa. Dra. Bernadette Farias Lóscio
Centro de Informática / UFPE
(Orientadora)

*Ao meu Deus, ao meu maravilhoso esposo Matheus, a meu pai David, minha mãe Ester,
minha irmã Rayssa, minha vó Bel, e aos amigos que contribuíram para a realização
deste trabalho.*

AGRADECIMENTOS

Tenho tanto a agradecer a Deus que nessas horas faltam as palavras para descrever tamanha gratidão. Mas sei que preciso começar pedindo perdão... Perdão por todas as vezes que pensei em desistir, por todas as vezes que eu não tive fé e forças suficientes para deixar a ansiedade de lado e confiar que tudo iria dar certo. Por todas as vezes que eu deixei o medo e a ansiedade falarem mais alto e, provavelmente, foi por causa disso que não consegui começar a escrever esses agradecimentos até que eu concluísse a última seção. Obrigada, Pai, por não ter me desamparado um minuto sequer, por ter me concedido capacidade mesmo sem eu merecer, mesmo com minhas falhas, foi graças à tua misericórdia e bondade que finalizei mais uma etapa da minha vida e sei que não seria possível sem tua ajuda.

Agradeço imensamente ao meu querido esposo Matheus, que está ao meu lado em mais esse momento. De forma alguma eu chegaria aqui sem você ao meu lado. Sem o teu abraço nos momentos difíceis em que o desespero tomou conta de mim. Não tenho como agradecer a Deus pelo presente que é você em minha vida. Eu não teria conseguido sem sua ajuda, obrigada por cuidar da casa esse período e fazer meu almoço sem reclamar (às vezes) todos os dias. Agora não tem mais desculpa, vai voltar a comer comida ruim feita por mim hehehe. Te amo!!!

Agradeço a toda a minha família que tem sido uma base forte pra mim e tem me ajudado em tudo que preciso, principalmente em orações. Obrigada painho e mainha por todas as palavras de consolo. Obrigada vovó Bel, que no momento que pensei em desistir devido à doença, a sra. me incentivou a não desistir e a lutar. A sra é o maior exemplo que eu poderia ter. Obrigada, Rayssa, minha irmã, que por diversas vezes me encorajou mesmo longe e me alegrou com um vídeo do bola mais safado do mundo que titia ama muito. Eu dedico essa conquista a vocês, que foram as principais razões para que eu concluísse esse mestrado.

Também não poderia deixar de agradecer a minha queridíssima orientadora Berna, por ter me orientado e me apoiado nesta jornada. Foram 5 meses intensos aturando minhas crises de ansiedade e me acalmando. A sra tem o dom de me acalmar, com suas palavras meigas e encorajadoras. Obrigada por cada ensinamento passado da graduação até aqui. Quando eu estava pronta para dizer que queria desistir, a sra começou a me falar palavras encorajadoras, sem nem saber que eu tinha chegado naquela reunião prestes a dar tchau. Meu muito obrigada por isso!!

Também não poderia deixar de agradecer a todos os meus amigos, em especial minha co-orientadora Karina que, mesmo sem o título oficial, esteve presente em reuniões, me ajudou nos momentos de desesperos via whats sempre me encorajando e se dispôs a me ajudar na elaboração do grupo focal. Meu muito obrigada de coração, só Deus pra te

recompensar.

Agradeço a Ruan, Sandrine, Glória e Wilker, sempre dispostos a compartilhar desesperos, rir de nervoso, e a estar presente para ajudar um ao outro. Vocês são demais, pessoal!!! E Ruan... Ainda dá tempo de fugir, mas não foge não hahaha. Agradeço também a toda equipe do grupo de pesquisa Aladin, pelas reuniões, discussões, confraternizações. Lairson, Marcelo, Helton, obrigada por toda ajuda ao longo desses dois anos de mestrado.

De forma alguma eu poderia esquecer da minha amiga Mirella. Mi, você é um presente de Deus pra mim. Obrigada por compartilhar momentos de desespero comigo. Nunca vou esquecer do dia em que eu estava em um ônibus lotado, em pleno trânsito, a ponto de ter uma crise de ansiedade quando recebi uma mensagem tua no Whats. Passamos, juntas, por muitos momentos de medo e angústia, mas que bom que ajudamos uma a outra a chegar até aqui. Que Deus fortaleça mais e mais nossa amizade!

Também não posso esquecer de agradecer ao meu digníssimo líder Vinicius e Fábio por me ajudarem nesses últimos meses. Vinicius por me liberar, sem nenhum problema, todas as vezes que pedi. Fábio por ter me ajudado nos dias de plantão em que precisei estar em reuniões na UFPE ou até mesmo para me cobrir enquanto eu estudava. Obrigada, meninos, vocês são 10!

Aos meus alunos da escolinha de música pelas orações e pela compreensão nesse período. Se preparem porque a moleza acabou, agora ninguém mais finge que estuda. Beca, obrigada por ter sido tão presente nesses últimos meses, por tua amizade, palavras e orações. Ao naipe mais lindo e a todos da orquestra que compreenderam minhas faltas e me ajudaram em oração.

Posso ter esquecido de algum nome, mas, mesmo sem citar, obrigada a todos que contribuíram para realização deste trabalho. Aos professores do CIn que contribuíram para meu crescimento acadêmico e profissional ao longo desses 7 anos na UFPE.

Por fim, agradeço à Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE) pelo financiamento que possibilitou a realização deste trabalho.

*“E, tudo o que pedirdes em oração, crendo, o recebereis. Mateus 21:22” (SAGRADA,
1994)*

RESUMO

Com o crescente interesse em utilizar a Web como principal meio de compartilhamento de dados e informações, práticas de publicação e consumo de dados na Web têm aumentado à medida em que a mesma passou a ser utilizada mais do que um simples meio de entretenimento. Além disso, muitas iniciativas foram propostas com o intuito de facilitar tanto a publicação quanto o consumo de dados, como metodologias, *guidelines* e boas práticas. Entretanto, muitos provedores de dados desconhecem ou não fazem uso destas iniciativas, não sendo capazes de enfrentar, de forma adequada, os desafios que surgem ao se publicar dados na Web, o que gera cada vez mais conjuntos de dados pobres em conteúdo, mesmo tendo um domínio de relevância. Embora consigamos achar na Web conjuntos de dados dos mais variados domínios, como dados governamentais, de saúde e de educação, não foram encontradas iniciativas que avaliem o processo de publicação de dados na Web de uma maneira geral. Diante deste cenário, este trabalho propõe um Modelo de Maturidade para Dados na Web (Data on the Web Maturity Model – DWMM) para avaliar a maturidade de organizações ou indivíduos no processo de publicação de dados na Web do ponto de vista técnico. O modelo foi avaliado através de um grupo focal, onde foi possível coletar dados qualitativos para evoluir o trabalho. O resultado desta avaliação norteou melhorias ao modelo proposto e, embora grande parte das sugestões tenham sido incluídas, uma nova avaliação será realizada a fim de considerar as sugestões que não foram atendidas para esta versão do modelo.

Palavras-chaves: Dados na Web. Publicação de Dados. Modelo de Maturidade. Ciclo de Vida. Boas Práticas.

ABSTRACT

With the increasing interest in using the Web as the primary means of data and information sharing, Web publishing and data consumption practices have increased as the Web has become more than just a means of entertainment. In addition, many initiatives have been proposed to facilitate both publication and data consumption, such as methodologies, guidelines and best practices. However, many data providers are unaware or does not make use of these initiatives, been unable to meet the challenges that arise when publishing data on the Web, which increasingly generates content-poor datasets, even with a domain of relevance. While we can find datasets on the Web from a wide range of domains, such as government, health and education data, no initiatives have been found to assess the process of publishing data on the Web in a general way. Given this scenario, this paper proposes a Data on the Web Maturity Model (DWMM) to assess the maturity of organizations or individuals in the process of publishing data on the Web from a technical point of view. The model was evaluated through a focus group, where it was possible to collect qualitative data to evolve the work. The result of this evaluation guided improvements to the proposed model and, although most of the suggestions were included, a new evaluation will be carried out in order to consider the suggestions that were not met for this version of the model.

Keywords: Data on the Web. Data Publishing. Maturity Model. Lifecycle. Best Practices.

LISTA DE FIGURAS

Figura 1 – Dataset - (%) População Pobre - 2010.	23
Figura 2 – Dados na Web x Dados Abertos x Dados Conectados.	24
Figura 3 – Modelo de Ciclo de Vida de Dados na Web - DWLM.	32
Figura 4 – Representação por Estágios do CMMI.	39
Figura 5 – Representação Contínua do CMMI.	40
Figura 6 – Visão Geral do <i>The Business Process Maturity Model</i>	42
Figura 7 – Processos de Gestão de Dados - <i>Open Data Maturity Model</i>	43
Figura 8 – Avaliação da Maturidade - <i>Open Data Pathway</i>	44
Figura 9 – <i>Open Government Maturity Model</i> - OGMM	45
Figura 10 – <i>Data Literacy Maturity Model</i> - DLMM.	46
Figura 11 – <i>Zakat Big Data Maturity Model</i>	47
Figura 12 – Níveis de Maturidade do <i>TDWI Big Data Maturity Model</i>	48
Figura 13 – Avaliação da Maturidade <i>TDWI Assessment</i>	49
Figura 14 – Componentes de um Modelo de Maturidade.	49
Figura 15 – Metamodelo do Modelo de Referência <i>MANGve Maturity Model</i> (M3).	50
Figura 16 – Etapas da Metodologia de Desenvolvimento do DWMM.	54
Figura 17 – Representação por Estágios do DWMM.	59
Figura 18 – Dimensões e Elementos do DWMM.	60
Figura 19 – Estrutura do DWMM.	75
Figura 20 – Avaliação dos Níveis do DWMM.	80
Figura 21 – Avaliação do Número de Níveis do DWMM.	81
Figura 22 – Avaliação do Número de Níveis do DWMM.	82
Figura 23 – Avaliação do Conjunto de Dimensões do DWMM.	83
Figura 24 – Avaliação da Dimensão Planejamento.	84
Figura 25 – Avaliação dos Objetivos da Dimensão Planejamento.	85
Figura 26 – Avaliação dos Níveis da Dimensão Planejamento.	85
Figura 27 – Avaliação da Dimensão Criação.	88
Figura 28 – Avaliação dos Objetivos da Dimensão Planejamento.	88
Figura 29 – Avaliação dos Níveis da Dimensão Planejamento.	89
Figura 30 – Avaliação da Dimensão Publicação.	90
Figura 31 – Avaliação dos Objetivos da Dimensão Publicação.	90
Figura 32 – Avaliação dos Níveis da Dimensão Publicação.	91
Figura 33 – Avaliação da Dimensão Consumo.	92
Figura 34 – Avaliação dos Objetivos da Dimensão Consumo.	93
Figura 35 – Avaliação dos Níveis da Dimensão Consumo.	93
Figura 36 – Avaliação da Dimensão Refinamento.	94

Figura 37 – Avaliação dos Objetivos da Dimensão Refinamento.	95
Figura 38 – Avaliação dos Níveis da Dimensão Refinamento.	95
Figura 39 – Avaliação da Dimensão Remoção.	96
Figura 40 – Avaliação dos Objetivos da Dimensão Remoção.	97
Figura 41 – Avaliação dos Níveis da Dimensão Remoção.	97

LISTA DE TABELAS

Tabela 3 – Benefícios x Boas Práticas dos Dados na Web	28
Tabela 6 – Representação Contínua x por Estágios	41
Tabela 29 – Informações dos Participantes do Grupo Focal	78
Tabela 31 – Perguntas Gerais sobre o DWMM	98

LISTA DE QUADROS

Quadro 1 – Desafios da publicação de dados na Web	26
Quadro 2 – Visão Geral do DWLM	35
Quadro 3 – Modelos de Maturidade na área de dados de uma forma geral	38
Quadro 4 – Comparação entre os modelos de maturidade encontrados na revisão ad-hoc da literatura	51
Quadro 5 – Características Gerais do DWMM	57
Quadro 6 – Características de Projeto do DWMM	57
Quadro 7 – Desafios da Dimensão Planejamento	61
Quadro 8 – Dimensão Planejamento do DWMM	62
Quadro 10 – Desafios da Dimensão Criação	63
Quadro 9 – Elementos da Dimensão Planejamento	64
Quadro 11 – Dimensão Criação do DWMM	65
Quadro 12 – Elementos da Dimensão Criação	65
Quadro 13 – Desafios da Dimensão Publicação	66
Quadro 14 – Dimensão Publicação do DWMM	67
Quadro 15 – Elementos da Dimensão Publicação	68
Quadro 16 – Desafios da Dimensão Consumo	69
Quadro 17 – Dimensão Consumo do DWMM	69
Quadro 18 – Elementos da Dimensão Consumo	70
Quadro 19 – Desafios da Dimensão Refinamento	71
Quadro 20 – Dimensão Refinamento do DWMM	72
Quadro 21 – Elementos da Dimensão Refinamento	73
Quadro 22 – Desafios da Dimensão Remoção	73
Quadro 23 – Dimensão Remoção do DWMM	74
Quadro 24 – Elementos da Dimensão Remoção	74

LISTA DE ABREVIATURAS E SIGLAS

ADLM	Abstract Data Lifecycle Model
DLMM	Data Literacy Maturity Model
DWLM	Data on the Web Lifecycle Model
DWMM	Data on the Web Maturity Model
ETL	Extract Transform Load
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
RDF	Resource Description Framework
URI	Uniform Resource Identifiers
W3C	World Wide Web Consortium
WWW	World Wide Web

SUMÁRIO

1	INTRODUÇÃO	17
1.1	MOTIVAÇÃO	17
1.2	CARACTERIZAÇÃO DO PROBLEMA	19
1.3	OBJETIVOS E CONTRIBUIÇÕES	20
1.4	ESTRUTURA DA DISSERTAÇÃO	21
2	VISÃO GERAL DOS DADOS NA WEB	22
2.1	VISÃO GERAL	22
2.1.1	Dados Abertos	23
2.1.2	Dados Conectados	24
2.2	BOAS PRÁTICAS PARA PUBLICAÇÃO DE DADOS NA WEB	25
2.3	DATA ON THE WEB LIFECYCLE MODEL (DWLM)	31
2.4	CONSIDERAÇÕES FINAIS	35
3	MODELOS DE MATURIDADE	36
3.1	VISÃO GERAL	36
3.2	MODELOS DE MATURIDADE	36
3.2.1	<i>Capability Maturity Model Integration - CMMI</i>	38
3.2.2	The Business Process Maturity Model - BPMM	41
3.2.3	Open Data Maturity Model	42
3.2.4	Open Government Maturity Model - OGMM	44
3.2.5	Data Literacy Maturity Model - DLMM	46
3.2.6	Zakat Big Data Maturity Model	47
3.2.7	TDWI Big Data Maturity Model	47
3.3	CONSIDERAÇÕES FINAIS	49
4	MODELO DE MATURIDADE PARA DADOS NA WEB	52
4.1	VISÃO GERAL DO DWMM	52
4.2	METODOLOGIA	53
4.2.1	Definição do Problema	54
4.2.2	Revisão dos Modelos de Maturidade Existentes	54
4.2.3	Projeto e Desenvolvimento do Modelo	55
4.2.4	Avaliação do Modelo	55
4.2.5	Comunicação	56
4.3	CARACTERÍSTICAS E CLASSIFICAÇÃO	56
4.3.1	Características Gerais	56

4.3.2	Características de Projeto do Modelo de Maturidade	57
4.4	DESCRIÇÃO DOS NÍVEIS DO DWMM	58
4.5	DIMENSÕES DO DWMM	58
4.5.1	Planejamento	61
4.5.2	Criação	63
4.5.3	Publicação	66
4.5.4	Consumo	68
4.5.5	Refinamento	70
4.5.6	Remoção	73
4.6	CONSIDERAÇÕES FINAIS	75
5	AVALIAÇÃO DO DWMM	76
5.1	GRUPO FOCAL	76
5.2	ORGANIZAÇÃO DO GRUPO FOCAL	76
5.2.1	Planejamento	76
5.2.2	Condução do Grupo Focal	78
5.2.3	Análise dos Dados	79
5.2.3.1	Avaliação dos Níveis de Maturidade	79
5.2.3.2	Avaliação do Conjunto de Dimensões do DWMM	82
5.2.3.3	Avaliação da Dimensão Planejamento	84
5.2.3.4	Avaliação da Dimensão Criação	87
5.2.3.5	Avaliação da Dimensão Publicação	90
5.2.3.6	Avaliação da Dimensão Consumo	92
5.2.3.7	Avaliação da Dimensão Refinamento	94
5.2.3.8	Avaliação da Dimensão Remoção	96
5.2.3.9	Perguntas Gerais	98
5.3	CONSIDERAÇÕES FINAIS	99
6	CONCLUSÃO	100
6.1	CONSIDERAÇÕES FINAIS	100
6.2	TRABALHOS FUTUROS	101
	REFERÊNCIAS	103
	APÊNDICE A – FORMULÁRIO DE AVALIAÇÃO DO DWMM	109

1 INTRODUÇÃO

Este Capítulo fornece uma visão geral desta pesquisa e apresenta o contexto no qual este trabalho está inserido. A Seção 1.1 apresenta uma breve motivação para o desenvolvimento deste trabalho. A caracterização do problema é descrita na Seção 1.2. Os objetivos e as principais contribuições deste trabalho são descritos na Seção 1.3 e, por fim, a Seção 1.4 apresenta a estrutura desta dissertação.

1.1 MOTIVAÇÃO

A World Wide Web (WWW) tornou-se um dos principais canais de compartilhamento e troca de informação. A quantidade de dados produzida e compartilhada ao longo da Web tem crescido rapidamente (BARNAGHI; SHETH; HENSON, 2013). Cerca de 57% da população mundial tem acesso à Internet nos dias atuais (WE ARE SOCIAL, 2019) e o potencial da Web como plataforma de compartilhamento e troca de informação tem sido intensificado tanto pelo grande volume de dados que são gerados por meio das mídias sociais e do paradigma de Web das Coisas (do inglês, *Web of Things*), quanto pelo crescimento da publicação online de dados abertos em todo o mundo (BARNAGHI; SHETH; HENSON, 2013; MADHAVAN et al., 2007).

Muitas organizações estão produzindo e disponibilizando uma grande quantidade de dados em uma variedade de plataformas (MADHAVAN et al., 2007). É importante notar que o interesse na publicação e compartilhamento de dados na Web não é algo novo (ABITEBOUL S.; SUCIU, 2000; BERNERS-LEE T.; SWICK, 1999). Porém, devido ao grande crescimento de dados disponibilizados e à flexibilidade oferecida pela Web, novos desafios precisam ser enfrentados para garantir o seu sucesso como plataforma de compartilhamento de dados, como por exemplo, lidar com o volume e a heterogeneidade dos dados disponíveis na Web (BARNAGHI; SHETH; HENSON, 2013).

Com essa grande e crescente quantidade de dados disponíveis na Web, podemos considerar os Ecossistemas de Dados na Web como sendo um conjunto de agentes envolvidos na produção, distribuição e consumo dos dados por meio da Web (LÓSCIO; GUIMARÃES; CALEGARI, 2016). Assim, dentre esses agentes temos os publicadores e consumidores de dados, sendo o publicador o ator responsável por compartilhar/disponibilizar os dados na Web, e o consumidor o ator que faz uso desses dados.

Em um mapeamento sistemático, Santos et al. (2018) investigou os benefícios relacionados à publicação e consumo de dados na Web e identificou 28 benefícios, dentre os quais podemos destacar os 6 benefícios mais citados nos 46 artigos analisados: melhor reuso de dados, fácil descoberta de dados, benefícios econômicos, interoperabilidade e benefícios sociais.

Todavia, para que esses benefícios possam ser alcançados ainda há vários problemas que precisam ser solucionados. Segundo Lóscio, Oliveira e Bittencourt (2015), a heterogeneidade dos dados e a falta de padrões para descrição e acesso aos conjuntos de dados tornam o processo de publicação, compartilhamento e consumo de dados uma tarefa complexa. Além disso, a falta de metadados e a incompletude nos dados afetam diretamente na sua qualidade (SANTOS et al., 2018).

Para solucionar tais problemas uma série de iniciativas foram propostas, desde metodologias (*e.g.* (NECASKÝ et al., 2013)), *guidelines* (*e.g.* (RADULOVIC et al., 2015)) e boas práticas (*e.g.* (LÓSCIO; BURLE; CALEGARI, 2016)) para publicação de dados a Sistemas para Gerenciamento de Dados na Web (*e.g.* (OLIVEIRA et al., 2018)). Entretanto, há uma série de aspectos e áreas importantes, identificados por Santos et al. (2018), que necessitam de pesquisa adicional e uma dessas áreas é a de Modelos de Maturidade.

Modelos de maturidade são amplamente utilizados em Engenharia Organizacional e de Software. De acordo com Fisher (2004), modelos de maturidade são usados para comparar e avaliar melhorias, permitindo assim que o grau de evolução em certos domínios seja medido. No ambiente de negócios, eles visam ajudar as organizações a identificar maneiras de melhorar a qualidade de seus processos e reduzir seu tempo de execução, proporcionando-lhes vantagens competitivas.

Dessa forma, alguns modelos de maturidade foram propostos para domínios como: Software (PAULK et al., 1993), Processos de Negócio (WEBER; CURTIS; GARDINER, 2008), dentre outros. No domínio de dados na Web, Dodds e Newman (2015) propuseram o *Open Data Maturity Model*, para avaliar as diversas maneiras nas quais a prática de dados abertos pode impactar uma organização. O modelo também fornece orientação às organizações sobre áreas potenciais para melhoria e possibilita comparar organizações para salientar seus respectivos pontos fortes e fracos, dando suporte a adoções mais amplas de melhores práticas e ajudando a aperfeiçoar processos. Entretanto, o modelo de (DODDS; NEWMAN, 2015) foca em aspectos organizacionais voltados à publicação de dados e reuso de dados; ou seja, não é possível avaliar a maturidade do processo de publicação dos dados na Web do ponto de vista técnico.

O processo de publicação dos dados na Web pode ser descrito a partir de um modelo de ciclo de vida. Um modelo de ciclo de vida¹ é utilizado para descrever um grupo de atividades e a forma como elas se relacionam. Desta forma, Silva (2019) propõe um Modelo de Ciclo de Vida de Dados na Web *Data on the Web Lifecycle Model* (DWLM) que descreve o conjunto de etapas que um dado ou conjunto de dados passa ao longo de sua vida na Web. O modelo foi proposto a fim de prover um entendimento comum do processo de publicação e consumo de dados entre os publicadores e os consumidores de dados.

Diante desse contexto, esta dissertação propõe um Modelo de Maturidade para Dados na Web (*Data on the Web Maturity Model* (DWMM)). Espera-se que com a utilização do

¹ https://pt.wikipedia.org/wiki/Modelo_de_ciclo_de_vida

DWMM seja possível avaliar a maturidade de organizações ou indivíduos no processo de publicação de dados na Web. Para isto consideramos as seguintes premissas: (i) o processo de publicação segue o Modelo de Ciclo de Vida dos Dados na Web (DWLM) proposto por (SILVA, 2019), (ii) são considerados apenas os aspectos técnicos da publicação de dados na Web, abordados de acordo com as Boas Práticas de Dados na Web (*Data on the Web Best Practices* - DWBP) (LÓSCIO; BURLE; CALEGARI, 2016), ou seja, aspectos políticos e organizacionais não são levados em consideração.

1.2 CARACTERIZAÇÃO DO PROBLEMA

Ao mesmo tempo que a flexibilidade da Web proporciona uma maior facilidade e comodidade para que organizações dos diversos tipos publiquem dados na Web, a complexidade em gerar valor a partir dos dados publicados aumenta na maioria das vezes. Em geral, os dados estão disponíveis de forma muito heterogênea e em formatos que dificultam o processamento automático desses dados por máquinas.

Oliveira et al. (2016), em um estudo realizado com os portais de dados abertos brasileiros, apontou que na maioria dos conjuntos de dados avaliados não foram disponibilizados metadados adequados ou suficientes para a descrição dos conjuntos de dados. Com isso, os consumidores de dados se prejudicam por não compreenderem o significado dos dados, sua estrutura e questões como termos de licença, qualidade dos dados, métodos de acesso e atualização dos conjuntos de dados (LÓSCIO; BURLE; CALEGARI, 2016). Também foi identificado que uma grande parcela dos conjuntos de dados foi publicada não abordando adequadamente as boas práticas para dados na Web (LÓSCIO; BURLE; CALEGARI, 2016).

Nesse contexto, constatamos que, embora as iniciativas de publicação de dados na Web, sejam em formato aberto e/ou conectado, tenham crescido, muitos provedores não fazem uso das Boas Práticas para Dados na Web propostas em (LÓSCIO; BURLE; CALEGARI, 2016) ou por desconhecimento ou por não saberem como aplicar e, conseqüentemente, não conseguem enfrentar adequadamente os desafios dos dados na Web; o que ocasiona na publicação de diversos conjuntos de dados que não são úteis para seus consumidores, por não prover uma fácil compreensão e manipulação para seus usuários finais.

Em geral, como a qualidade dos dados é um fator crucial para o consumo e compartilhamento de dados, é importante que haja uma avaliação prévia não apenas dos dados, mas do conjunto de dados como um todo (SANTOS et al., 2018). Nesse contexto, os modelos de maturidade podem ser aplicados para fornecer recomendações claras sobre como conduzir melhorias no processo de publicação de dados a fim de gerar conjuntos de dados úteis para seus consumidores finais.

Um modelo de maturidade na área de dados na Web será útil para auxiliar organizações e indivíduos que almejam a publicação de dados com qualidade, seguindo os princípios das boas práticas, a fim de garantir que esses conjuntos de dados sejam facilmente reutilizados, compreendidos, processados e de fato úteis para seus consumidores.

Geralmente, os modelos de maturidade avaliam processos bem definidos dentro de uma organização e possuem estruturas bastante formais além de descrições de processos complexos. No caso dos dados na Web, não existe um processo consolidado de publicação de dados na Web. Dessa forma, a fim de guiar a avaliação da maturidade no processo de publicação de dados na Web, escolhemos o Modelo de Ciclo de Vida dos Dados na Web, DWLM, como as dimensões do modelo proposto neste trabalho, uma vez que possui estrutura adaptável, flexível e de fácil compreensão.

No entanto, para avaliar de fato o processo de publicação de dados na Web é necessário, ainda, avaliar os desafios que surgem durante as diferentes fases do DWLM. Só então, poderemos analisar como as boas práticas têm sido aplicadas para enfrentar os desafios dos dados na Web a fim de proporcionar a criação de conjuntos de dados úteis, além de viabilizar benefícios aos publicadores e consumidores de dados.

1.3 OBJETIVOS E CONTRIBUIÇÕES

O principal objetivo desta dissertação é propor um Modelo de Maturidade para Dados na Web. Esse modelo leva em consideração o conjunto de fases do DWLM, avaliando como os desafios dos dados na Web têm sido enfrentados em cada fase do ciclo de vida dos dados na Web. Espera-se que com o tratamento adequado de cada um dos desafios o grau de utilidade dos conjuntos de dados aumente; em outras palavras, entendemos que a utilidade de um conjunto de dados pode ser medida a partir dos benefícios alcançados ao se aplicar as Boas Práticas para Dados na Web.

Para alcançar o objetivo geral acima, foram definidos os seguintes objetivos específicos:

- *Levantamento dos principais Modelos de Maturidade existentes:* Foi feita uma revisão da literatura de forma *ad-hoc* para que pudéssemos ter um ponto de partida para elaboração do modelo proposto. Desta forma, analisamos tanto modelos nas áreas de Engenharia de Software e Processos, como modelos voltados à área de Dados Abertos e Governo Aberto;
- *Definição de um Modelo de Maturidade para Dados na Web:* Após o levantamento dos principais modelos de maturidade, foi possível propor o modelo proposto neste trabalho para avaliar organizações e indivíduos durante o processo de publicação de dados na Web por meio da avaliação das fases do Modelo de Ciclo de Vida dos Dados na Web;
- *Avaliação do modelo proposto por meio de um grupo focal:* Por fim, foi realizada uma sessão de grupo focal para avaliar a viabilidade e adequação do modelo proposto.

1.4 ESTRUTURA DA DISSERTAÇÃO

Os próximos capítulos estão organizados como se segue. No Capítulo 2 é apresentada uma visão geral dos dados na Web. No Capítulo 3 são apresentados alguns modelos de maturidade encontrados na literatura. Já no Capítulo 4, é apresentado o modelo de maturidade para dados na Web proposto nesta dissertação. No Capítulo 5 apresentamos a técnica de grupo focal utilizada para avaliação do modelo. Por fim, no Capítulo 6 é feita uma breve discussão sobre o trabalho realizado e sugestões para os trabalhos futuros.

2 VISÃO GERAL DOS DADOS NA WEB

Este Capítulo, apresenta os conceitos que se fazem necessários para o entendimento do domínio da área de Dados na Web. Inicialmente, na Seção 2.1, é apresentada uma visão geral do contexto de Dados na Web. Na Seção 2.2 detalhamos as Boas Práticas para Publicação de Dados na Web e os desafios enfrentados na publicação e consumo dos dados na Web. Na Seção 2.3 é introduzido o conceito de Ciclo de Vida para Dados e, em seguida, discutido o Modelo de Ciclo de Vida de Dados na Web proposto por Silva (2019) o qual é utilizado como base para definição deste trabalho. E, por fim, na Seção 2.4 são apresentadas as Considerações Finais do Capítulo.

2.1 VISÃO GERAL

O interesse na utilização da Web como plataforma para publicação de dados não é algo novo (BERNERS-LEE T.; SWICK, 1999; ABITEBOUL S.; SUCIU, 2000). Porém, nos últimos anos este interesse tem aumentado devido à flexibilidade que a Web provê para publicação e consumo de dados. Tanto o governo como organizações não governamentais disponibilizam uma variedade de dados na Web, abrangendo domínios como Educação, Economia, Segurança e Dados Científicos.

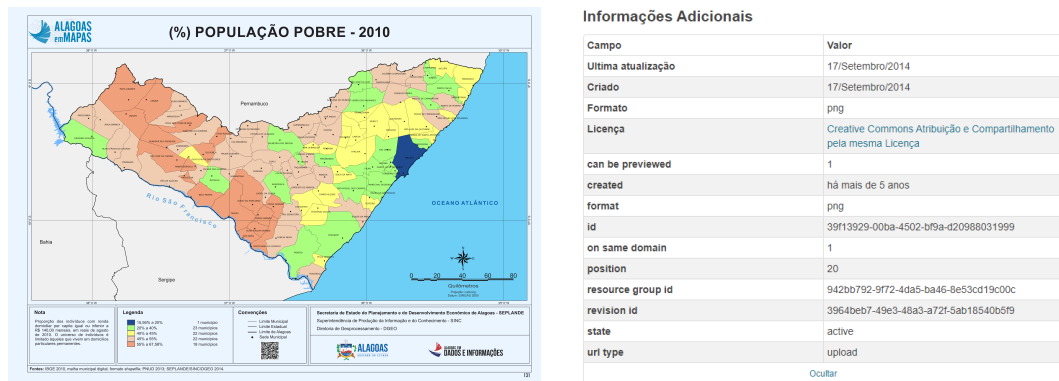
Com o cenário de crescimento na publicação de dados na Web, diversos benefícios têm sido gerados para a sociedade e inúmeras aplicações buscam fazer uso desses conjuntos de dados, com o objetivo de gerar informações úteis e relevantes e até mesmo a geração de novos dados. Com essa expansão no volume de dados é importante observar aspectos relacionados à proveniência e à qualidade dos dados.

Muitos conjuntos de dados disponibilizados na Web têm como característica a ausência completa ou parcial de um esquema que defina rigorosamente a estrutura dos dados a serem armazenados e muitos dados não são fáceis de serem consumidos, devido à complexidade ou a falta de informações disponíveis, como metadados que permitam descrever a proveniência dos dados, a frequência e o histórico de atualizações, o que acaba não passando confiabilidade ao consumidor.

Em uma breve consulta ao portal de dados abertos do estado de Alagoas, encontramos o dataset com o título “(%) População Pobre - 2010” com a proporção dos indivíduos com renda domiciliar per capita igual ou inferior a R\$ 140,00 mensais, em reais de agosto de 2010. O primeiro problema que encontramos ao analisar este dataset se dá pelo formato dos dados que consiste em uma única imagem no formato .png (ver Figura 1). Além disso, o dataset possui poucos metadados descritivos, com campos onde não há uma definição prévia de seus significados. O dataset também não possui uma frequência de atualização e sua última modificação foi a data de criação do mesmo, o que acaba dificultando seu

consumo.

Figura 1 – Dataset - (%) População Pobre - 2010.



Fonte: Extraído de <<http://dados.al.gov.br/>>

Segundo Lóscio, Guimarães e Calegari (2016), Dados na Web é um termo mais geral que pode ser usado para denotar dados publicados de acordo com os princípios da arquitetura da Web¹. Como ilustrado na Figura 2, Dados na Web podem ser classificados como Dados Abertos (*Open Data*) (PIRES, 2015), Dados Conectados (*Linked Data*) e Dados Abertos Conectados (*Linked Open Data*) (BERNERS-LEE, 2006). Lóscio et al. (2018) acrescenta que é importante notar que nem todos os dados disponíveis na Web são compartilhados abertamente. Em outras palavras, os provedores de dados determinam a política de publicação de dados e em quais circunstâncias os dados devem ser publicados.

2.1.1 Dados Abertos

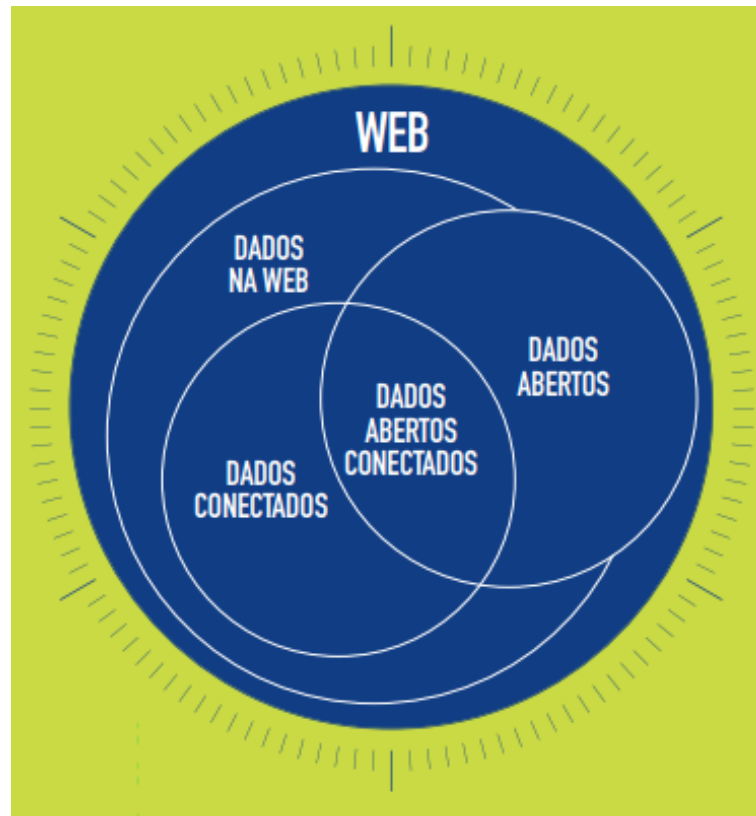
O conceito de Dados Abertos aplica-se a todo dado publicado na Web disponível para que qualquer usuário possa utilizar, reutilizar e redistribuir esse dado sem qualquer restrição de patentes, propriedade intelectual ou outro mecanismo de controle, estando sujeito, no máximo, a atribuição de autoria (DIETRICH et al., 2009).

A publicação de dados em formato aberto tem gerado benefícios em diversas áreas, como a transparência dos órgãos governamentais, que provê melhor acesso aos dados, participação e colaboração da sociedade em seus governos (MANUAL, 2011). A abertura dos dados também pode contribuir para o avanço da ciência, bem como a geração de novos empregos; pode aumentar, também, a qualidade nos serviços prestados por diversas organizações, além de muitos outros benefícios que podem ser encontrados em (PIRES, 2015).

Segundo (ISOTANI; BITTENCOURT, 2015), é possível que diferentes organizações e sistemas possam trabalhar de forma colaborativa; para tal, os dados devem seguir três normas fundamentais propostas por (DIETRICH et al., 2009):

¹ <https://www.w3.org/TR/webarch/>

Figura 2 – Dados na Web x Dados Abertos x Dados Conectados.



Fonte: (LÓSCIO et al., 2018)

1. **Disponibilidade e Acesso:** Os dados devem estar disponíveis como um todo e sob custo razoável de reprodução, preferencialmente possíveis de serem baixados pela Internet e em formato conveniente e modificável;
2. **Reuso e redistribuição:** Os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados, além de garantir a legibilidade por máquinas;
3. **Participação universal:** Todos devem ser capazes de usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos.

2.1.2 Dados Conectados

Dados Conectados (*Linked Data*) diz respeito a um conjunto de princípios para publicação de dados estruturados na Web, com o intuito de criar uma “Web de Dados” (BIZER; HEATH; BERNERS-LEE, 2009). Um dos principais objetivos do *Linked Data* é prover uma Web onde os dados possam estar diretamente ligados com outros dados por meio de *links* Resource Description Framework (RDF)², possibilitando a navegação entre diferentes

² Um modelo de representação de dados e/ou objetos (recursos) que podem ser identificados na Web através de URIs.

conjuntos de dados e permitindo a realização de inferências.

Para Bizer, Heath e Berners-Lee (2009) os Dados Conectados baseiam-se em duas tecnologias que são fundamentais para a Web: o Uniform Resource Identifiers (URI), identificador único de recursos na Web, e o Hypertext Transfer Protocol (HTTP)³, protocolo de aplicação utilizado para requisições na Web.

Os princípios de *Linked Data* foram introduzidos por Berners-Lee (2006) e resumem-se em quatro princípios básicos:

- I. Utilização de URIs como nomes para recursos;
- II. Utilização de URIs HTTP para que pessoas consigam encontrar esses nomes;
- III. Quando a URI for acessada, garantir que informações úteis possam ser obtidas por meio dessa URI, as quais devem estar representadas no formato RDF e SPARQL;
- IV. Incluir *links* para outras URIs de forma que outros recursos possam ser descobertos.

2.2 BOAS PRÁTICAS PARA PUBLICAÇÃO DE DADOS NA WEB

Com o crescente interesse na publicação e consumo de dados na Web, Lóscio, Oliveira e Bittencourt (2015) observaram que nos últimos anos, a heterogeneidade dos dados e a falta de padrões para descrição e acesso aos conjuntos de dados tornaram o processo de publicação, compartilhamento e consumo de dados uma tarefa complexa. Em termos gerais, os publicadores de dados visam compartilhar dados abertamente ou com acesso controlado, enquanto os consumidores de dados buscam ser capazes de encontrar, usar e estabelecer conexões entre os dados. Isso cria uma necessidade fundamental para um entendimento comum entre os provedores e consumidores de dados (LÓSCIO; GUIMARÃES; CALEGARI, 2016).

Na busca por esse entendimento comum, um grupo de trabalho do World Wide Web Consortium (W3C) intitulado de *Data on the Web Best Practices Working Group* propôs uma série de recomendações desenvolvidas para incentivar e permitir a expansão continuada da Web, bem como fornecer orientação técnica para a publicação de dados na Web, contribuindo para melhorar a relação entre provedores e consumidores de dados (LÓSCIO et al., 2018; LÓSCIO; BURLE; CALEGARI, 2016). O documento conta com 35 boas práticas que oferecem orientação técnica para a publicação de dados na Web.

A fim de determinar o escopo das boas práticas, um conjunto de casos de uso⁴ foi compilado, onde cada caso de uso fornece um cenário de como os dados são comumente publicados e como eles são usados na Web. A partir desses casos de uso, foi possível identificar os principais desafios enfrentados pelos produtores e consumidores de dados.

³ <https://www.w3.org/Protocols/rfc2616/rfc2616.html>

⁴ <https://www.w3.org/TR/dwbp-ucr/>

De uma forma geral, os desafios dos dados na Web envolvem a resolução tanto de questões técnicas, como seleção de formato de dados usado, quanto não técnicas, como licenciamento e engajamento dos usuários (Ver Quadro 1) (LÓSCIO; OLIVEIRA; BITTENCOURT, 2015). Assim como as boas práticas, os desafios foram identificados a partir dos casos de uso propostos pelo *Data on the Web Working Group*.

Quadro 1 – Desafios da publicação de dados na Web

Desafio	Descrição
Metadados	Permitir que os seres humanos entendam os metadados, interpretando a natureza e a estrutura dos dados, e que as máquinas também possam processá-los
Licença	Permitir que os seres humanos compreendam as informações da licença e que as máquinas possam detectar automaticamente
Proveniência	Permitir que os seres humanos conheçam a origem ou o histórico do conjunto de dados e que as máquinas possam processar automaticamente tais informações
Qualidade	Documentar a qualidade dos dados, para facilitar o processo de seleção dos conjuntos de dados e chances de reutilização
Versionamento	Permitir que versões dos dados sejam geradas e seja possível o acesso a cada versão
Identificação	Fornecer identificadores únicos para os conjuntos de dados e distribuições
Formato	Escolher formatos que permitam o uso e o reuso
Vocabulários	A fim de melhorar a interoperabilidade e manter terminologia comum entre os produtores e consumidores
Acesso	Permitir o fácil acesso aos dados usando a infraestrutura da Web tanto para seres humanos quanto para máquinas
Preservação	A fim de indicar corretamente se os dados foram removidos ou arquivados
Feedback	Receber feedback dos consumidores e assegurar que os dados atendam as suas necessidades
Enriquecimento	Enriquecer, melhorar ou refinar os dados brutos agregando valor
Republicação	Permitir que os dados utilizados possam ser republicados

Fonte: Derilinx, Lóscio e Archer (2015)

Os desafios surgem como uma série de questões que devem e podem ser respondidas por meio das boas práticas para dados na Web. Uma delas é “Como prover metadados para humanos e máquinas?”, para responder essa questão ou desafio, três boas práticas podem ser aplicadas. A primeira delas é “BP1 - Prover metadados”, a segunda “BP2 - Prover metadados descritivos” e por último a “BP3 - Prover metadados estruturais”.

Conforme descrito por (LÓSCIO; BURLE; CALEGARI, 2016), cada boa prática tem um Resultado esperado, que, em geral, é uma melhoria no modo que um consumidor de dados (humano ou software) pode manipular um conjunto de dados publicados na Web ou, em alguns casos, uma melhoria no próprio conjunto de dados, o que também resultará em um ganho para o consumidor de dados.

Desta forma, as boas práticas ajudam o publicador de dados a enfrentar os desafios que surgem ao se publicar e consumir dados na Web, mostrando exemplos de uso das boas

práticas e descrevendo os benefícios que o publicador receberá devido à adoção das boas práticas durante a produção do conjunto de dados.

Os benefícios identificados por Lóscio, Guimarães e Calegari (2016) ao se aplicar as boas práticas são oito, onde cada benefício representa uma melhoria na forma em que os conjuntos de dados estão disponíveis na Web. Cada benefício está relacionado a uma ou mais boas práticas. Lóscio, Guimarães e Calegari (2016) notam que o benefício se torna mais forte na medida em que aumenta a adoção das boas práticas. Esses benefícios são listados a seguir:

- *Compreensão*: diz respeito a um melhor entendimento sobre a estrutura, significado dos dados, metadados e natureza do conjunto de dados por parte das pessoas. Este benefício se relaciona com 10 boas práticas.
- *Processabilidade*: diz respeito a capacidade das máquinas em processar e manipular dados dentro de um conjunto de dados automaticamente. Este benefício, por sua vez, está relacionado com 9 boas práticas.
- *Descoberta*: diz respeito à facilidade que os agentes de software terão em descobrir automaticamente um conjunto de dados ou dados dentro de um conjunto de dados, caso o publicador adote as boas práticas relacionadas. Ao todo, este benefício se relaciona com 7 boas práticas.
- *Reuso*: este benefício é relacionado a todas as 35 boas práticas e diz respeito à reutilização do conjunto de dados por diferentes grupos de consumidores de dados.
- *Confiabilidade*: aplicando as 19 boas práticas relacionadas, o conjunto de dados passará confiança a seus consumidores.
- *Conectividade*: adotando as 4 boas práticas relacionadas, será possível criar *links* entre recursos e conjuntos de dados.
- *Acesso*: diz respeito à capacidade em que os humanos e máquinas terão em acessar dados atualizados em uma variedade de formas aplicando as 8 boas práticas relacionadas.
- *Interoperabilidade*: diz respeito à facilidade de chegar a um consenso entre publicadores e consumidores de dados, utilizando as 8 boas práticas relacionadas.

A Tabela 3 relaciona os 8 benefícios às 35 boas práticas dos dados na Web.

Tabela 3 – Benefícios x Boas Práticas dos Dados na Web

Benefícios	Boas Práticas
Acesso	<i>BP17 - Fornecer “bulk download”</i>
	<i>BP18 - Fornecer subconjuntos para conjuntos de dados grandes</i>
	<i>BP19 - Usar “negociação de conteúdo” para servir os dados disponíveis em vários formatos</i>
	<i>BP20 - Fornecer acesso em tempo real</i>
	<i>BP21 - Fornecer dados atualizados</i>
	<i>BP23 - Tornar os dados disponíveis por meio de uma API</i>
	<i>BP24 - Usar padrões Web como base para construção de APIs</i>
Compreensão	<i>BP32 - Fornecer visualizações complementares</i>
	<i>BP1 - Fornecer metadados</i>
	<i>BP2 - Fornecer metadados descritivos</i>
	<i>BP3 - Fornecer metadados estruturais</i>
	<i>BP5 - Fornecer informações de proveniência dos dados</i>
	<i>BP13 - Usar representações de dados que sejam independentes de localidade (Locale Neutral)</i>
	<i>BP15 - Reutilizar vocabulários, dando preferência aos padronizados</i>
	<i>BP16 - Escolher o nível de formalização adequado</i>
	<i>BP29 - Coletar feedback dos consumidores de dados</i>
Conectividade	<i>BP31 - Enriquecer dados por meio da geração de novos dados</i>
	<i>BP32 - Fornecer visualizações complementares</i>
	<i>BP9 - Usar URIs persistentes como identificadores de conjuntos de dados</i>
	<i>BP10 - Usar URIs persistentes como identificadores dentro de conjuntos de dados</i>
	<i>BP18 - Fornecer subconjuntos para conjuntos de dados grandes</i>
	<i>BP24 - Usar padrões Web como base para construção de APIs</i>

Continua na próxima página

Tabela 3 – Continuação

Benefícios	Boas Práticas
Confiabilidade	<i>BP4 - Fornecer informações sobre a licença dos dados</i>
	<i>BP5 - Fornecer informações de proveniência dos dados</i>
	<i>BP6 - Fornecer informação de qualidade de dados</i>
	<i>BP7 - Fornecer indicador de versão</i>
	<i>BP8 - Fornecer o histórico de versões</i>
	<i>BP11 - Atribuir URIs para as versões dos conjuntos de dados e séries</i>
	<i>BP15 - Reutilizar vocabulários, dando preferência aos padronizados</i>
	<i>BP22 - Fornecer uma explicação para os dados que não estão disponíveis</i>
	<i>BP25 - Fornecer documentação completa para as APIs</i>
	<i>BP26 - Evitar alterações que afetem o funcionamento de sua API</i>
	<i>BP27 - Preservar identificadores</i>
	<i>BP28 - Avaliar a cobertura do conjunto de dados</i>
	<i>BP29 - Coletar feedback dos consumidores de dados</i>
	<i>BP30 - Compartilhar o feedback disponível</i>
	<i>BP31 - Enriquecer dados por meio da geração de novos dados</i>
	<i>BP32 - Fornecer visualizações complementares</i>
	<i>BP33 - Fornecer feedback para o provedor original</i>
	<i>BP34 - Obedecer os termos de licença</i>
	<i>BP35 - Citar a publicação original do conjunto de dados</i>
Continua na próxima página	

Tabela 3 – Continuação

Benefícios	Boas Práticas
Descoberta	<i>BP1 - Fornecer metadados</i>
	<i>BP2 - Fornecer metadados descritivos</i>
	<i>BP9 - Usar URIs persistentes como identificadores de conjuntos de dados</i>
	<i>BP10 - Usar URIs persistentes como identificadores dentro de conjuntos de dados</i>
	<i>BP11 - Atribuir URIs para as versões dos conjuntos de dados e séries</i>
	<i>BP24 - Usar padrões Web como base para construção de APIs</i>
Interoperabilidade	<i>BP35 - Citar a publicação original do conjunto de dados</i>
	<i>BP9 - Usar URIs persistentes como identificadores de conjuntos de dados</i>
	<i>BP10 - Usar URIs persistentes como identificadores dentro de conjuntos de dados</i>
	<i>BP15 - Reutilizar vocabulários, dando preferência aos padronizados</i>
	<i>BP16 - Escolher o nível de formalização adequado</i>
	<i>BP23 - Tornar os dados disponíveis por meio de uma API</i>
	<i>BP24 - Usar padrões Web como base para construção de APIs</i>
	<i>BP26 - Evitar alterações que afetem o funcionamento de sua API</i>
	<i>BP33 - Fornecer feedback para o provedor original</i>
Continua na próxima página	

Tabela 3 – Continuação

Benefícios	Boas Práticas
Processabilidade	<i>BP1 - Fornecer metadados</i>
	<i>BP3 - Fornecer metadados estruturais</i>
	<i>BP12 - Usar formatos de dados padronizados legíveis por máquina</i>
	<i>BP14 - Fornecer dados em vários formatos</i>
	<i>BP15 - Reutilizar vocabulários, dando preferência aos padronizados</i>
	<i>BP18 - Fornecer subconjuntos para conjuntos de dados grandes</i>
	<i>BP23 - Tornar os dados disponíveis por meio de uma API</i>
	<i>BP24 - Usar padrões Web como base para construção de APIs</i>
	<i>BP31 - Enriquecer dados por meio da geração de novos dados</i>
Reuso	<i>Todas as boas práticas</i>

Fonte: Adaptado de (LÓSCIO; BURLE; CALEGARI, 2016)

2.3 DATA ON THE WEB LIFECYCLE MODEL (DWLM)

Um ciclo de vida, de uma maneira geral, pode ser definido por uma série de transformações que um ser vivo, um objeto ou um artefato pode passar durante sua vida. No contexto de dados não é diferente; o ciclo de vida dos dados na Web é o processo de gerenciamento de informações desde a seleção dos dados a serem publicados até o consumo e remoção do acesso a esses dados. Todo dado disponível na Web passa ou já passou por um conjunto de fases que compõe o processo de publicação e consumo dos dados (MÖLLER, 2013).

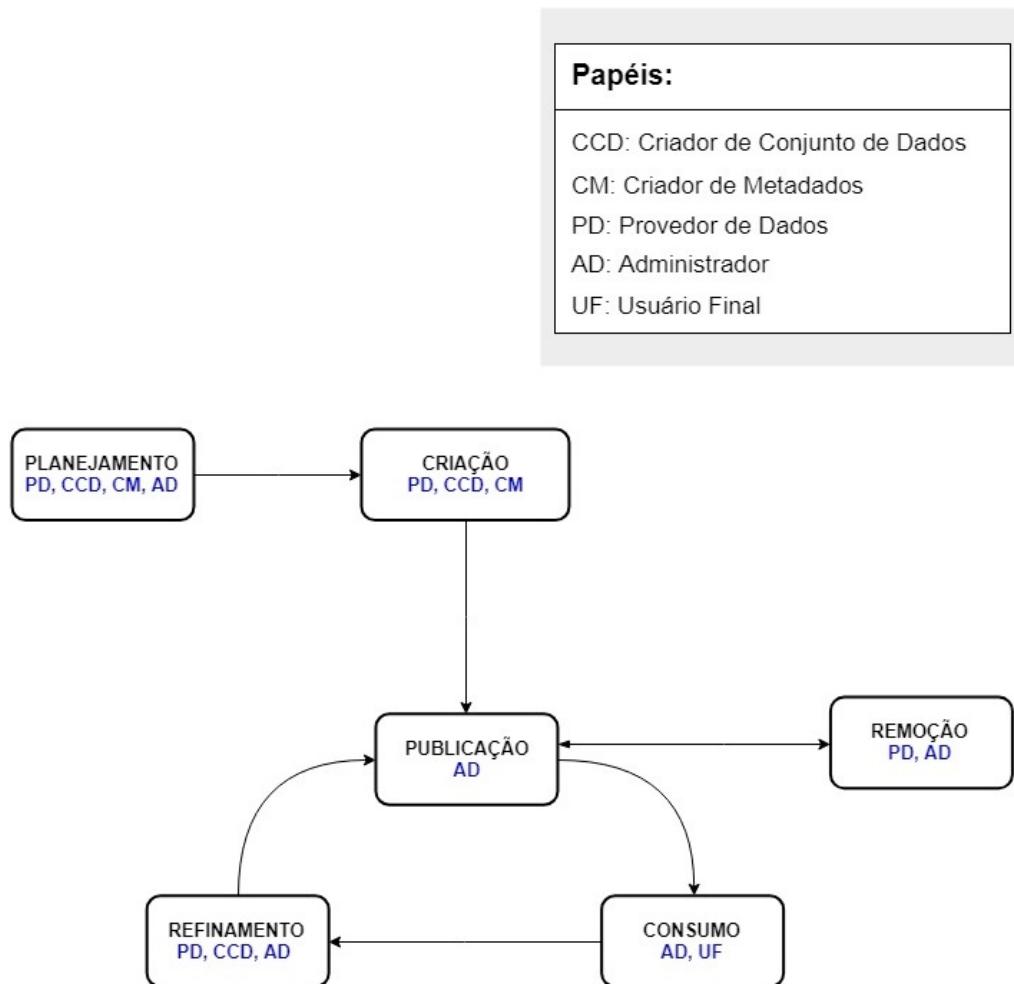
Silva (2019) propõe um Modelo de Ciclo de Vida de Dados na Web chamado de DWLM com o objetivo de prover um entendimento comum das etapas que um conjunto de dados passa ao longo de sua vida na Web, além de assegurar que os conjuntos de dados publicados atendam a alguns requisitos que permitam seu processamento por humanos e máquinas.

Embora encontremos alguns ciclos de vida no contexto de dados na literatura (*e.g.* (CHEN; CHEN; LIN, 2003; CATTEAU; VIDAL; BROISIN, 2006; MÖLLER, 2013; ALSHAMMARI; SIMPSON, 2017; LÓSCIO; OLIVEIRA; BITTENCOURT, 2015)), eles geralmente estão relacionados a domínios mais específicos ou não contemplam de forma abrangente o domínio de Dados na Web (SILVA, 2019). Desta forma, escolhemos o DWLM como base para a definição do DWMM.

O DWLM foi baseado no Abstract Data Lifecycle Model (ADLM) proposto por Möller (2013) e o Ciclo de Vida de Dados na Web proposto por Lóscio, Oliveira e Bittencourt

(2015). Silva (2019) ressalta ainda que o DWLM foi feito para tratar um conjunto de dados a cada iteração do conjunto de fases (ver Figura 3).

Figura 3 – Modelo de Ciclo de Vida de Dados na Web - DWLM.



Fonte: (SILVA, 2019)

O modelo de Silva (2019) possui 5 papéis que atuam ao longo de 6 fases com 16 atividades distribuídas entre elas:

- Planejamento

1. *Especificar fontes de dados*

Diz respeito à especificação das fontes de dados que poderão ser usadas para a coleta dos dados que serão publicados. Também são definidas as estratégias de coleta para cada fonte especificada;

2. *Descrever o conjunto de dados*

Diz respeito à coleta de informações relevantes para a compreensão e a criação do conjunto de dados, como dados de proveniência, metadados descritivos e estruturais, a licença dos dados e os formatos de distribuição;

3. *Estabelecer solução para publicação do conjunto de dados*

A partir das informações já contidas no *Documento de Descrição do Conjunto de Dados*, será possível ter uma visão geral do propósito do conjunto de dados e, dessa forma, definir a solução de publicação. Uma solução é classificada como *primitiva* se apenas disponibiliza os dados por meio de uma página HyperText Markup Language (HTML) com, no mínimo, a opção de *download*. Uma solução é classificada como *básica* se estende as funcionalidades da *primitiva* e disponibiliza os dados e seus metadados numa página HTML em formato legível por máquina. Uma solução é classificada como *intermediária* se estende as funcionalidades da *básica* e oferece o acesso ao conjunto de dados e seus metadados por meio de uma API. Por fim, uma solução é classificada como *avançada* se estende as funcionalidades da *intermediária* e oferece pesquisas (*e.g.* filtragens no conjunto de dados, buscas por palavras-chaves) além de oferecer pré-visualizações e um ambiente para coleta de *feedback* dentre outras coisas.

- Criação

1. *Criar o conjunto de dados*

Diz respeito à criação do conjunto de dados a partir da extração dos dados conforme a necessidade de cada fonte. Essa extração pode ocorrer a partir da (i) disponibilização de um *link* para que os dados provenientes de APIs em tempo real possam ser acessados via Web ou (ii) realização de um processo de Extract Transform Load (ETL) para dados que necessitam passar por processos de transformação, integração ou limpeza.

2. *Avaliar qualidade*

Diz respeito à escolha das métricas de qualidade que o *Criador de Conjunto de Dados* julgar importantes e avaliar a qualidade do seu conjunto de dados.

3. *Validar o conjunto de dados recém criado*

Antes da publicação dos dados, o *Provedor de Dados* verificará se o conjunto de dados corresponde às expectativas dele.

- Publicação

1. *Publicar o conjunto de dados de acordo com a solução escolhida*

Nesta atividade será feita a publicação do conjunto de dados a partir da solução escolhida pelo *Administrador* na fase de *Planejamento*

2. *Tornar o conjunto de dados acessível*

Diz respeito à definição de padrões URLs para acesso ao conjunto de dados e a disponibilização do conjunto fazendo uso dos protocolos padrões de acesso da Web (i.e HTTP).

3. *Fornecer alternativas de uso*

Diz respeito à disponibilização de visualizações complementares (*i.e* gráficos e exemplos de uso do conjunto de dados) para que os consumidores possam compreender facilmente os dados.

- Consumo

1. *Acessar o conjunto de dados*

Diz respeito ao acesso ao conjunto de dados por meio dos consumidores.

2. *Fazer uso do conjunto de dados*

A depender da licença escolhida para o conjunto de dados, os consumidores poderão optar por usá-lo para a realização de atividades adicionais.

3. *Prover e disponibilizar feedback*

Essa atividade diz respeito à disponibilização de um ambiente colaborativo entre publicadores e consumidores para coleta e disponibilização de *feedback*.

- Refinamento

1. *Corrigir e enriquecer o conjunto de dados*

Consiste na correção do conjunto de dados a partir de um processo de limpeza, objetivando o aumento da qualidade dos dados, por meio da detecção e remoção de erros e inconsistência. O enriquecimento dos dados diz respeito à agregação de valor aos conjuntos de dados tanto por meio da adição de novos dados e metadados, quanto por anotações semânticas.

2. *Validar o conjunto de dados refinado*

Após a correção realizada na atividade anterior, é recomendado a criação de um *log* que especifique as alterações de forma que o *Provedor de Dados* possa validar.

3. *Versionar o conjunto de dados*

Diz respeito à nova versão do conjunto de dados gerada a partir das atividades anteriores.

- Remoção

1. *Remover acesso ao conjunto de dados*

Diz respeito ao processo de preservação do identificador do conjunto de dados, relacionado a BP27. Corresponde a todo o processo de informar ao consumidor dos dados o motivo da remoção do conjunto de dados e à avaliação de cobertura do conjunto, caso ele possua distribuições em RDF.

Uma visão geral do DWLM pode ser vista no Quadro 2. Nele são descritas as fases do DWLM, bem como um breve resumo do que cada fase representa e as atividades que acontecem em cada fase.

Quadro 2 – Visão Geral do DWLM

Fase	Descrição	Atividades
<i>Planejamento</i>	Fase em que ocorre a descrição do conjunto de dados, bem como a definição dos metadados que serão disponibilizados juntamente com o conjunto. Além disso, informações de proveniência, licença e volume do conjunto de dados são consideradas nessa fase	Especificar fontes de dados; Descrever o conjunto de dados; Estabelecer solução para publicação do conjunto de dados.
<i>Criação</i>	Fase em que ocorre a criação dos dados propriamente dita e onde serão definidas as estratégias de coleta para cada uma das fontes de dados de origem descritas no Documento de Descrição do Conjunto de Dados.	Criar o conjunto de dados; Avaliar qualidade; Validar o conjunto de dados recém criado.
<i>Publicação</i>	A fase onde o conjunto de dados e seus metadados deverão ser disponibilizados na Web de acordo com a solução escolhida para sua publicação	Publicar o conjunto de dados de acordo com a solução escolhida; Tornar o conjunto de dados acessível; Fornecer alternativas de uso.
<i>Consumo</i>	Representa as diferentes formas de uso e manipulação dos dados publicados.	Acessar conjunto de dados; Fazer uso do conjunto de dados; Prover e disponibilizar feedback;
<i>Refinamento</i>	Onde ocorre operações de identificação e correção de erros, adição e atualização de dados, metadados e semântica.	Corrigir e enriquecer o conjunto de dados; Validar o conjunto de dados refinado; Versionar o conjunto de dados.
<i>Remoção</i>	Onde ocorre a realização da preservação do conjunto de dados que terá o seu acesso removido.	Remover acesso ao conjunto de dados.

Fonte: A autora (2019)

2.4 CONSIDERAÇÕES FINAIS

Neste Capítulo, foi apresentada uma visão geral dos Dados na Web, descrevendo o contexto ao qual a publicação de dados está envolvida. Em seguida, apresentamos o conceito de Dados Abertos e Dados Conectados descrevendo suas respectivas características. Na Seção 2.2 apresentamos as Boas Práticas para publicação de dados na Web e os benefícios gerados a partir da adoção das mesmas. Em seguida, apresentamos os Desafios dos dados na Web. Por fim, finalizamos o Capítulo descrevendo o Modelo de Ciclo de Vida dos Dados na Web (DWLM) que será usado como principal componente do Modelo proposto neste trabalho.

3 MODELOS DE MATURIDADE

Este Capítulo apresenta o conceito de Modelo de Maturidade. Na Seção 3.1 apresentamos uma visão geral do conceito de maturidade a partir da definição de diversos autores. Na Seção 3.2 descrevemos as principais características de um modelo de maturidade e, em suas subseções são mostrados alguns exemplos de Modelos de Maturidade encontrados na literatura. Por fim, na Seção 3.3 são apresentadas as Considerações Finais do Capítulo.

3.1 VISÃO GERAL

A literatura apresenta diversos conceitos de maturidade, todos relacionados com estágios de desenvolvimento. Paulk et al. (1993) descrevem maturidade como um processo específico para definir, gerenciar, medir e controlar explicitamente o crescimento evolutivo de uma entidade. Anderson e Jessen (2003) definem maturidade como um estado no qual uma organização é perfeitamente capaz de atingir objetivos que ela própria estabelece. Fitterer e Rohner (2010) sugere que a maturidade está associada a um critério de avaliação ou ao estado de ser completo, perfeito e pronto. E Sen, Ramamurthy e Sinha (2011) descrevem maturidade como um conceito que progride de um estado inicial para um estado final (que é mais avançado), isto é, níveis mais altos de maturidade. Similarmente, Mettler (2009) fala que a maturidade está relacionada com o progresso evolucionário em demonstrar uma capacidade particular ou a busca de um certo objetivo, de um estado inicial para um estado desejável final. Seja qual for o ponto de vista, podemos dizer que a maturidade está relacionada com a evolução tanto de um indivíduo como de uma organização como um todo.

A mensuração da maturidade pode ser realizada por meio de modelos de maturidade. Alguns autores definem modelos de maturidade como instrumentos que permitem mensurar o grau de evolução em determinados domínios e contribuem sistematicamente para direcionar o desenvolvimento, posicionamento e as transformações organizacionais (BRUIN et al., 2005; PAULK et al., 1993; BECKER; KNACKSTEDT; PÖPPELBUSS, 2009).

3.2 MODELOS DE MATURIDADE

Embora a utilização de modelos de maturidade não seja algo tão novo assim (NOLAN, 1973), os modelos de maturidade passaram a ser amplamente disseminados a partir da criação do *Capability Maturity Model* (CMM) (PAULK et al., 1993). Existe, atualmente, uma variedade de modelos de maturidade, muitos deles fundamentados no CMM, com o objetivo de apoiar e dirigir as estratégias de gestão de projetos, processos de software, gestão de conhecimento, dentre outras áreas.

Mettler e Rohner (2009) argumentam que a compreensão de modelos de maturidade está situada entre modelos e métodos. Para eles, modelos refletem um estado de um determinado domínio de aplicação, seja a descrição exata da situação atual ou uma sugestão para um estado de destino mais eficiente ou ideal. Ao contrário dos modelos em que a reprodução das descrições de estado da solução ideal é o objetivo final (o quê), os métodos focam na especificação de atividades para alcançar a solução ideal (como).

Nesse sentido, modelos de maturidade podem ser distinguidos em sua natureza como descrito a seguir (BRUIN et al., 2005; STERNKOPF; MUELLER, 2018; BECKER et al., 2010; PROENÇA; BORBINHA, 2018):

1. *Descritivos*: reportam o nível atual que uma entidade se encontra. Não apresenta formas de melhorias da maturidade e não tem relações com o desempenho. É utilizado para avaliar o presente, a real situação da organização.
2. *Prescritivos*: indicam a forma para alcançar os próximos níveis de maturidade, detalhando atividades específicas de melhorias.
3. *Comparativos*: permitem a realização de benchmarking em organizações ou regiões, realizando analogias. São aplicados com o objetivo de obter referência de maturidade dentro de diferentes entidades.

Becker et al. (2010) ainda ressalta que modelos podem assumir natureza Descritiva e Prescritiva ao mesmo tempo. Quanto à tipologia dos modelos de maturidade Fraser, Moultrie e Gregory (2002) divide em 3 grupos básicos:

1. *Maturity Grids*: contém textos descritivos para cada atividade em cada nível de maturidade; é de complexidade moderada, requerendo no máximo algumas páginas de texto.
2. *CMM-Like*: possui arquitetura particular que é mais formal (e complexa), onde cada área de processo é organizada por características comuns que especificam um número de práticas chave para abordar uma série de objetivos.
3. *Likert-Like Questionnaires*: quando construído de uma maneira particular, pode ser considerado uma forma simples de modelo de maturidade.

Na revisão *ad-hoc* da literatura encontramos alguns modelos de maturidade já disseminados no mercado, na área de processos e engenharia de software, como também alguns modelos no contexto de dados de uma maneira geral e voltados ao contexto da Web, dentre os quais podemos citar os que encontram-se no Quadro 3.

Quadro 3 – Modelos de Maturidade na área de dados de uma forma geral

Título	Referência
<i>Maturity Model for Data Literacy in Non-governmental Organizations</i>	(STERNKOPF; MUELLER, 2018)
<i>Maturity Model for Corporate Data Quality Management</i>	(HÜNER; OFNER; OTTO, 2009)
<i>IDC's Big Data and Analytics MaturityScape</i>	(VESSET et al., 2015)
<i>TDWI Big Data Maturity Model Guide</i>	(HALPER; KRISHNAN, 2013)
<i>E-Government Maturity Model for Sustainable E-Government Services from the Perspective of Developing Countries</i>	(JOSHI; ISLAM, 2018)
<i>A Model of Data Warehousing Process Maturity</i>	(SEN; RAMAMURTHY; SINHA, 2011)
<i>An Open Government Maturity Model for social media-based public engagement</i>	(LEE; KWAK, 2012)
<i>Open Data Maturity Model</i>	(DODDS; NEWMAN, 2015)

Fonte: A autora (2019)

Uma vez que nem todos os modelos encontrados dispunha de uma literatura que descrevesse mais claramente seus níveis e processos, nas subseções a seguir detalharemos apenas alguns dos modelos encontrados durante nossa revisão *ad-hoc* da literatura.

3.2.1 **Capability Maturity Model Integration - CMMI**

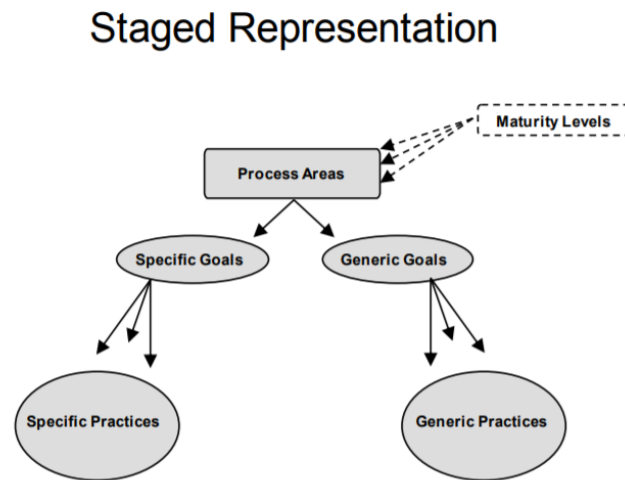
O *Capability Maturity Model Integration* - CMMI foi criado a partir - e para suprir as limitações - do modelo CMM (PAULK et al., 1993). Um dos grandes problemas do CMM foi a demanda por um grande esforço de integração quando áreas distintas da mesma organização precisavam ser avaliadas sob a mesma perspectiva de maturidade preconizada pelo CMM (NETO, 2015). Desta forma, o CMMI surgiu a fim de integrar várias áreas incluindo o gerenciamento de projetos, a engenharia de software, a engenharia de sistemas, de integração de produtos e o desenvolvimento de processos (SEI, 2010).

O objetivo do CMMI é servir de guia para a melhoria de processos na organização, assim como para auxiliar a habilidade dos profissionais em gerenciar o desenvolvimento de aquisição e manutenção de produtos ou serviços de software, além de proporcionar a visibilidade apropriada do processo de desenvolvimento para todos os envolvidos no projeto. Para tal, o modelo contempla duas representações divididas em níveis, priorizando as ações a serem realizadas de forma clara a partir de um caminho de melhoria evolutiva de processos *ad-hoc*, imaturos para processos disciplinados e maduros, com qualidade e eficácia aprimoradas (SEI, 2010; VASCONCELOS; MORAIS, 2010).

A primeira representação do CMMI, também presente no CMM, é a representação por estágios (ver Figura 4). Essa representação permite que as organizações melhorem um conjunto de processos inter-relacionados e, de forma incremental, tratem sucessivos

conjuntos de áreas de processo, representando a ideia de maturidade. Para atingir um nível particular, uma organização precisa satisfazer todos objetivos ou metas para determinada área de processo ou um conjunto de áreas de processos selecionadas para melhoria (SEI, 2010).

Figura 4 – Representação por Estágios do CMMI.



Fonte: Extraído de (SEI, 2010)

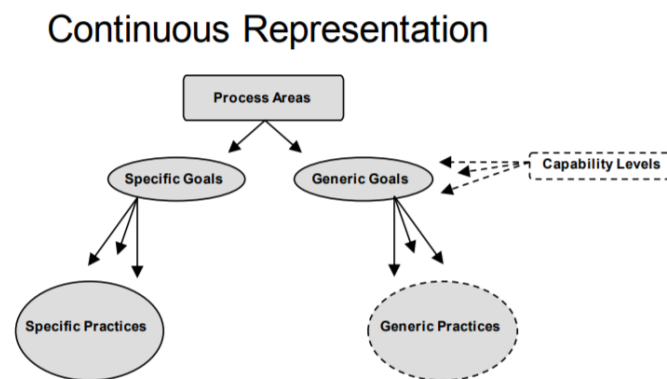
Nessa representação, as áreas de processo são organizadas pelos níveis de maturidade descritos a seguir (VASCONCELOS; MORAIS, 2010):

- **Nível 1 - Inicial:** Neste nível de maturidade não há áreas de processos e os processos realizados são normalmente imprevisíveis, caóticos e pobremente controlados.
- **Nível 2 - Gerenciado:** Neste nível, os projetos da organização têm a garantia de que os requisitos são gerenciados, planejados, executados, medidos e controlados. O gerenciamento de projetos é o foco principal deste nível.
- **Nível 3 - Definido:** Neste nível, os processos são bem mais caracterizados e entendidos e são descritos em padrões, procedimentos, ferramentas e métodos. O foco neste nível é a padronização do processo.
- **Nível 4 - Quantitativamente Gerenciado:** Os objetivos específicos e genéricos atribuídos para os níveis de maturidade 2, 3 e 4 foram alcançados e os processos são medidos e controlados.
- **Nível 5 - Em Otimização:** É o nível mais alto de maturidade CMMI, onde uma organização atinge todos os objetivos específicos atribuído para os níveis anteriores. Os processos são continuamente aperfeiçoados, considerando que a variação de um processo está relacionada às interações entre seus componentes tendo como foco principal a melhoria contínua do processo.

Ao contrário da representação por estágios, a representação contínua (ver Figura 5) permite que uma mesma organização atinja níveis de maturidade distintos em diferentes processos organizacionais (VASCONCELOS; MORAIS, 2010). Nesta representação, o foco é avaliar a maturidade de processos específicos. Esta representação possui 4 níveis de capacidade descritos a seguir (SEI, 2010):

- **Nível 0 - Incompleto:** Um processo é parcialmente realizado ou não, onde um ou mais objetivos específicos do processo não são satisfeitos.
- **Nível 1 - Realizado:** Um processo realizado satisfaz todos os objetivos específicos da área de processo e produz algum trabalho.
- **Nível 2 - Gerenciado:** É caracterizado como um processo realizado, planejado e executado de acordo com políticas pré-definidas.
- **Nível 3 - Definido:** É caracterizado como um processo que é personalizado a partir do conjunto de processos padrão da organização de acordo com as diretrizes de adaptação da organização. Indica claramente a finalidade, entradas, critérios de entrada, atividades, funções, medidas, etapas de verificação, saídas e critérios de saída.

Figura 5 – Representação Contínua do CMMI.



Fonte: Extraído de (SEI, 2010)

O que difere as duas representações é que a contínua se concentra na capacidade da área de processo medida pelos níveis de capacidade e a representação por estágios se concentra na maturidade geral medida pelos níveis de maturidade. Desta forma, Neto (2015) afirma que o foco da representação por estágios está na maturidade da organização, a qual é medida através de níveis de maturidade avaliados segundo um conjunto de múltiplas áreas de processo. O foco da representação contínua está em áreas de processo individuais que

são medidos através de seus níveis de capacidade. O Quadro 6 apresenta uma comparação entre os níveis de capacidade e de maturidade em ambos os tipos de representação.

Tabela 6 – Representação Contínua x por Estágios

Nível	Representação contínua (níveis de capacidade)	Representação por estágios (níveis de maturidade)
0	Incompleto	
1	Executado	Inicial
2	Gerenciado	Gerenciado
3	Definido	Definido
4		Quantitativamente Gerenciado
5		Em Otimização

Fonte: Traduzido de (SEI, 2010)

3.2.2 The Business Process Maturity Model - BPMM

Fisher (2004) propõe um modelo de maturidade multidimensional e não linear para avaliar os processos de negócio de uma organização. A primeira dimensão do modelo é composta de 5 componentes que ele intitula de “*Five Levers of Change*” ou “cinco alavancas de mudança” que são: *Strategy, Controls, Process, People* e *IT*.

A partir da avaliação das capacidades de uma determinada organização e do avanço delas, a organização pode progredir para a segunda dimensão proposta por Fisher (2004), na qual ele define como “*States of Process Maturity*” ou “estados de maturidade do processo” que podem ser: *Siloed, Tactically Integrated, Process Driven, Optimized Enterprise* e *Intelligent Operating Network*.

Fisher (2004) argumenta que o propósito do modelo é fazer com que as organizações identifiquem suas próprias lacunas de onde estão atualmente em cada uma das cinco alavancas de mudança e, desta forma, identificar quais ações específicas tomar para superar as limitações atuais e alcançar benefícios. Uma visão geral do modelo pode ser vista na Figura 6.

Por fim, Fisher (2004) pondera que as empresas raramente se encontram em um estado consistente em todas as cinco alavancas de mudança e que isso impede que uma organização alcance os benefícios de estar em um estado particular.

Não foi encontrado na literatura maiores descrições de como o modelo poderia ser aplicado na prática e conforme descrito por (NETO, 2015) muitas vezes o formalismo em excesso acaba dificultando a aplicabilidade do modelo em si. Apesar da área de Processos de Negócio ser vasta e com processos bem definidos, acreditamos que a multidimensionalidade do modelo de (FISHER, 2004) acaba tornando-o complexo de ser aplicado.

Figura 6 – Visão Geral do *The Business Process Maturity Model*.

	Isolado	Taticamente Integrado	Orientado a Processos	Empresa Otimizada	Rede Operacional Inteligente
Estratégia	- Reage às condições do mercado dentro de 1 a 2 anos, normalmente perseguindo um concorrente; - Integração dentro de funções; - Impulsionado por custo e eficiência.	- Adapta/reage à dinâmica do mercado em 12 meses; - Alguma integração multifuncional para resolver dores; - Entrada inicial na integração ponto a ponto com parceiros.	- Adapta/reage à dinâmica do mercado dentro de 3-6 meses; - Liderança de processo em toda a empresa é estabelecida; - O processo de negócios é o elemento fundamental da empresa.	- Adaptável à dinâmica do mercado dentro de semanas; - Empresa completamente organizada em torno de processos; - Processos otimizados + execução produzem vantagens competitivas.	- Capacidades preditivas e liderança de mercado; - Continuamente adaptável à dinâmica do mercado quase em tempo real; - A empresa e seus parceiros são organizados em torno de processos; - A vantagem competitiva é impulsionada e compartilhada pelos parceiros.
Controles	- Autoridade / autonomia local e funcional; - Sem padrões ou governança em toda a empresa; - Nenhum programa formal de medição de valor.	- Estrutura hierárquica de gerenciamento; - Decisões independentes do departamento funcional; - Padrões ou governança em toda a empresa limitados.	- A liderança formal do processo estabelece prioridades; - Business cases conduzem projetos; - Métricas de processo vinculadas ao desempenho individual e da equipe.	- Equipes de processos responsáveis pelo desempenho geral; - Métricas de processo relevantes institucionalizadas como principais medidas de desempenho.	- Equipes de processos entre empresas possuem desempenho próprio; - Métricas de processo relevantes são usadas para medir o desempenho bidirecional do parceiro.
Processo	- Processos de negócios estáticos; - Unidades funcionais; - Unidades geográficas; - Focado em departamentos; - Comunicações informais dentro dos departamentos.	- Reengenharia limitada de processos e coordenação interfuncional / de processos (geralmente manuais, esforços únicos); - Os sistemas conduzem as definições de processo da linha de base.	- Transição total do foco funcional para o processo, incluindo estrutura de gerenciamento, equipes de execução e avaliação de desempenho; - BPO direcionado.	- Integração total do processo em toda a empresa; - Compromisso com o programa de melhoria contínua de processos; - Terceiriza processos de negócios não essenciais (reduzir custos e melhorar a qualidade)	- Integração total do processo no ecossistema; - Os processos principais fluem perfeitamente entre os firewalls.
Pessoas	- Especialistas no assunto; - Cultura é contraditória, desconfiança mútua; - Nenhum procedimento formal de gerenciamento de mudanças; - "Eu farei o meu trabalho, você faz o seu".	- Membros multifuncionais / equipe de processos (geralmente liderados por TI); - Compreensão limitada das necessidades e dependências do processo interdepartamental.	- Líderes de processo definem, implantam, aprimoram e mantêm processos principais; - As equipes funcionais se concentram na execução de alta qualidade.	- Organização enxuta focada na otimização de definições e execução de processos; - Treinamento contínuo de processos para funcionários.	- A seleção de parceiros inclui atributos culturais e do processo; - Treinamento contínuo de processos para funcionários e parceiros.
TI	- Sistemas independentes; - Ilhas de automação; - Integração apenas dentro de funções; - Sistema(s) corporativo(s) legado(s).	- Alavancar sistemas ERP para integração interfuncional; - Integração de parceiros ponto a ponto; - A TI lidera iniciativas multifuncionais (focado em sistemas).	- TI suporta a equipe de liderança de processos em iniciativas; - Consolidação de sistemas e instâncias para otimizar processos e informações gerenciais.	- Utiliza soluções de gerenciamento de processos de negócios (BPM) para automatizar a execução, o monitoramento e o controle de processos em toda a Empresa.	- Utiliza soluções de Gerenciamento de processos de negócios (BPM) para automatizar e monitorar a execução de processos em todo o ecossistema.

Fonte: Traduzido de (FISHER, 2004)

3.2.3 Open Data Maturity Model

Diferentes Modelos de Maturidade voltados à temática de dados abertos estão sendo discutidos e propostos na literatura por diversos autores. A grande maioria desses modelos, encontrados durante uma revisão *ad-hoc* da literatura, têm sido propostos com foco em iniciativas de dados abertos, principalmente, em agências governamentais.

O Modelo de Maturidade de Dados Abertos proposto por Dodds e Newman (2015) foi desenvolvido como um projeto conjunto entre o Instituto de Dados Abertos (ODI)¹ e o Departamento de Meio Ambiente, Alimentação e Assuntos Rurais (Defra)². O modelo surgiu a partir dos resultados obtidos a partir de uma série de *workshops* com a participação de administradores de dados e especialistas em dados abertos do governo do Reino Unido e da comunidade de dados abertos em geral.

Segundo Dodds e Newman (2015), o modelo foi projetado para priorizar especificamente o modo como a prática de dados abertos impacta uma organização e consiste de 15 atividades organizacionais, como processos de liberação de dados ou adoção de normas

¹ <http://theodi.org>

² <https://www.gov.uk/government/organisations/departament-for-environment-food-rural-affairs>

comunitárias, que estão agrupadas em 5 temas (processos de gestão de dados, conhecimento e competências, suporte ao cliente envolvido, investimento e desempenho financeiro e supervisão estratégica).

O modelo de Dodds e Newman (2015) é avaliado a partir de 5 níveis de maturidade, baseado nos níveis do CMMI: 1 - Inicial, 2 - Reprodutível, 3 - Definido, 4 - Gerenciado e 5 - Otimizado, onde cada um dos níveis tem uma definição que descreve as principais características da atividade que ocorre naquele nível. Para níveis mais baixos, as atividades tendem a ser *ad-hoc*, enquanto em processos de níveis mais altos, serão mais refinadas. As atividades ainda, estão associadas a uma descrição dos efeitos benéficos que se seguem ao aumento da maturidade na atividade.

A estrutura do Guia do Modelo de Maturidade de Dados Abertos de Dodds e Newman (2015), pode ser vista brevemente na Figura 7.

Figura 7 – Processos de Gestão de Dados - *Open Data Maturity Model*.

Atividade	Aspecto	Nível 1 Inicial	Nível 2 Reprodutível	Nível 3 Definido	Nível 4 Gerenciado	Nível 5 Otimização	Efeitos benéficos
Processo de liberação de dados	Publicação	Poucos dados abertos publicados ou nenhum dado aberto publicado. Conjuntos de dados que são publicados (...)	Projetos ou produtos específicos podem ter definido um processo reprodutível para (...)	Há um processo de liberação padrão reprodutível em toda a organização para (...)	Todos os conjuntos de dados são liberados de acordo com o processo organizacional padrão	A organização coleta e monitora as métricas em seu processo de liberação, (...)	Redução de despesas gerais associadas a liberações de dados.

Fonte: Extraído de (DODDS; NEWMAN, 2015)

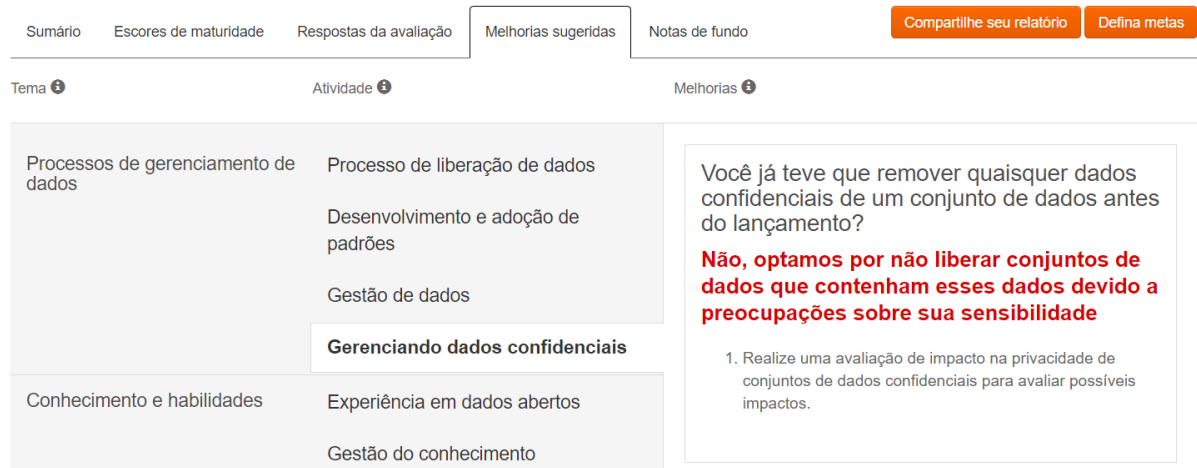
O modelo não tem como objetivo oferecer uma descrição prescritiva de como exatamente as organizações devem publicar ou reutilizar dados abertos. Em lugar disto, o modelo prioriza comportamentos generalizados que uma organização deve ter e os processos que deve adotar. Uma vez que organizações individuais devem implantar os processos definidos pelo modelo com base em suas necessidades, de acordo com os autores, seria excessivamente prescritivo recomendar processos que pudessem ser aplicados de forma útil em todas as organizações.

Uma avaliação completa com base neste modelo dará à organização uma pontuação de maturidade em uma série de atividades, a saber, como um dado é liberado, como é governado e como conjuntos de dados são avaliados. A pontuação irá refletir a maturidade dos processos da organização em uma área específica e pode ser usada para identificar áreas para melhorias e definição de metas mensuráveis (DODDS; NEWMAN, 2015).

A avaliação pode ser feita pela ferramenta disponibilizada pela ODI, *Open Data*

*Pathway*³. Um exemplo do resultado da avaliação da maturidade de uma organização hipotética pode ser visto na Figura 8.

Figura 8 – Avaliação da Maturidade - *Open Data Pathway*.



Fonte: Extraído de <<http://pathway.theodi.org>>

Para esta organização, foi avaliado um nível baixo de maturidade na execução da atividade “*Gerenciando dados confidenciais*”, desta forma foi sugerido a condução de uma avaliação de impacto de privacidade de conjuntos de dados sensíveis para avaliar possíveis impactos.

Em comparação aos outros modelos já mencionados, a estrutura do ODMM é mais simples de ser compreendida. No entanto, o modelo é voltado especificamente para aspectos organizacionais, do ponto de vista da publicação e do reuso dos dados em uma organização e sua ferramenta de avaliação fornece mais *insights* para obtenção da maturidade.

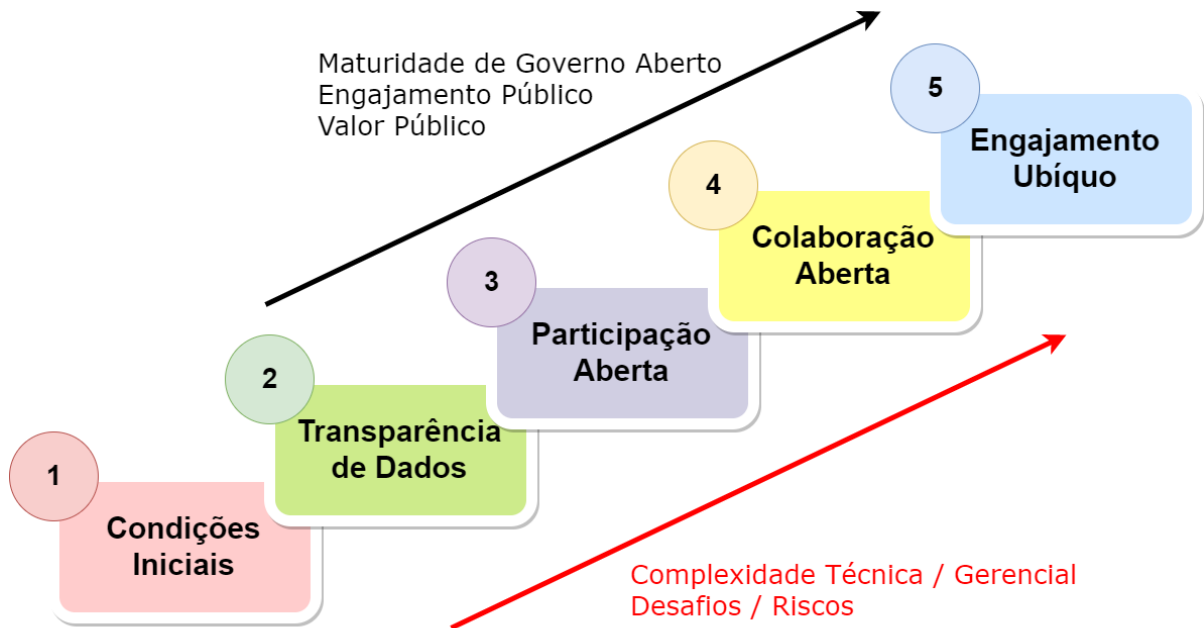
3.2.4 Open Government Maturity Model - OGMM

Lee e Kwak (2012) propõem um modelo de maturidade para avaliar e guiar iniciativas de governo aberto com foco em transparência, participação, colaboração e engajamento público por meio de mídias sociais. O modelo consiste em 5 níveis de maturidade: 1 - Condições iniciais, 2 - Transparência de Dados, 3 - Participação aberta, 4 - Colaboração aberta, 5 - Engajamento ubíquo. Os autores argumentam que há uma sequência lógica para aumentar o engajamento público baseado nas mídias sociais e as entidades governamentais devem se concentrar em alcançar um nível de maturidade de cada vez.

À medida que as agências governamentais avançam para um nível de maturidade mais elevado, o público está mais envolvido e, assim, é agregado mais valor à iniciativa de governo aberto. Por outro lado, um maior nível de maturidade enfrenta maior complexidade técnica e gerencial e maiores desafios e riscos (ver Figura 9).

³ <http://pathway.theodi.org/>

Figura 9 – *Open Government Maturity Model - OGMM*



Fonte: Extraído de (LEE; KWAK, 2012)

A utilização do *Open Government Maturity Model* (OGMM) ajuda as entidades governamentais a implementar suas iniciativas de governo aberto de maneira eficaz, construindo capacidades organizacionais e tecnológicas de maneira ordenada. Para cada nível de maturidade há uma série de desafios e melhores práticas que podem ser aplicadas (LEE; KWAK, 2012).

Lee e Kwak (2012) acreditam que, embora o modelo tenha sido desenvolvido principalmente com base em dados de agências federais dos Estados Unidos, ele pode ser utilizado por agências governamentais estaduais ou locais sem grandes modificações, pois os princípios e objetivos fundamentais do governo aberto são os mesmos para governos federais, estaduais e locais.

Embora não tenhamos encontrado outras literaturas com maiores detalhes da aplicabilidade do OGMM, em um único artigo, os autores conseguiram demonstrar o OGMM com clareza para seus possíveis usuários.

O modelo é composto basicamente por duas Tabelas. A primeira, para cada nível de maturidade é listado os focos (que entendemos serem áreas gerais dentro de uma organização governamental), as capacidades ou processos referentes a dados em si e a participação ou colaboração dos cidadãos, as principais saídas para cada nível de maturidade e as métricas utilizadas naquele nível.

Já a segunda Tabela, é demonstrado cada desafio que pode ser enfrentado em determinado nível de maturidade e as boas práticas que podem ser utilizadas para mitigar tais desafios.

3.2.5 Data Literacy Maturity Model - DLMM

*Data Literacy*⁴ ou alfabetização de dados é a capacidade de ler, entender, criar e comunicar dados como informação útil. Para Gray, Chambers e Bounegru (2012 apud STERNKOPF; MUELLER, 2018) a alfabetização de dados é a capacidade de consumir conhecimento, produzir coerentemente e pensar criticamente sobre dados.

Nesse contexto, Sternkopf e Mueller (2018) propõem um modelo de maturidade para alfabetização de dados (Data Literacy Maturity Model (DLMM)), desenvolvido no contexto de organizações não governamentais. O objetivo do DLMM é auxiliar a definir e avaliar os recursos de manuseio e conversão de dados, bem como identificar os pontos fortes e fracos das organizações sem fins lucrativos e fornecer orientação para futuras práticas de dados.

O modelo consiste de 11 dimensões: 1 - *data culture*, 2 - *data ethics and security*, 3 - *ask questions and define*, 4 - *find*, 5 - *get*, 6 - *verify*, 7 - *clean*, 8 - *analyze*, 9 - *visualize*, 10 - *communicate* e 11 - *assess and interpret* que, segundo os autores, foram definidas a partir de estudos mais recentes em *data literacy*. Essas dimensões são avaliadas ao longo de 4 níveis de maturidade: 1 - *uncertainty*, 2 - *enlightenment*, 3 - *certainty* e 4 - *data fluency* (ver Figura 10).

Figura 10 – *Data Literacy Maturity Model* - DLMM.

1 Incerteza	2 Iluminação	3 Certeza	4 Fluência de Dados
As organizações desconhecem a necessidade de habilidades de alfabetização de dados e não tem ou tem uma compreensão muito vaga do que é necessário. Os indivíduos podem ter um certo interesse em dados e trabalhar digitalmente, mas não têm certeza sobre as diferentes etapas existentes ao trabalhar com dados.	As organizações estão experimentando a aplicação de tópicos relacionados a dados. Descreve um estado em que muitos dados já foram entendidos teoricamente, mas não podem ser aplicados em muitos casos e precisam ser treinados.	As organizações executam etapas de tratamento de dados com confiança e têm atividades orientadas a dados incorporadas em seus processos de rotina, onde quer que isso faça sentido. Procedimentos e padrões genéricos sobre como lidar com dados são entendidos em todos os níveis da organização.	As organizações estabeleceram uma cultura de informação de dados em todos os níveis. Os dados são usados ativamente para melhorar processos e criar fluxos de trabalho.
Cultura de Dados			
Os dados são percebidos como um termo ambíguo que causa inseguranças.	Os dados são percebidos como um conceito interessante e os benefícios são apreciados. Existem inseguranças em relação aos casos de uso e o que exatamente esperar.	Os dados não são percebidos como fonte de insegurança, mas entendidos como facilitadores do progresso e do apoio às atividades existentes e planejadas. Líderes e gerentes superiores apoiam iniciativas de dados.	As barreiras psicológicas dos dados foram derrubadas (por exemplo, inseguranças, medo, resignação) e o conforto em torno dos dados é promovido. Gerentes de nível superior e gerentes de projeto entendem e apoiam a importância de recursos dedicados (tempo, orçamento, recursos humanos) para manipulação e conversão de dados.

Fonte: Traduzido de (STERNKOPF; MUELLER, 2018)

O DLMM fornece apenas uma visão geral de cada uma das 11 dimensões nos 4 níveis de maturidade e auxilia as organizações a identificar e analisar como as pessoas estão lidando com os dados e quais áreas precisam ser exploradas a fim de treinar os profissionais de forma mais intensiva nas competências de dados. Não encontramos literatura adicional que demonstre o modelo com mais clareza, no entanto achamos sua natureza descritiva fácil de ser compreendida.

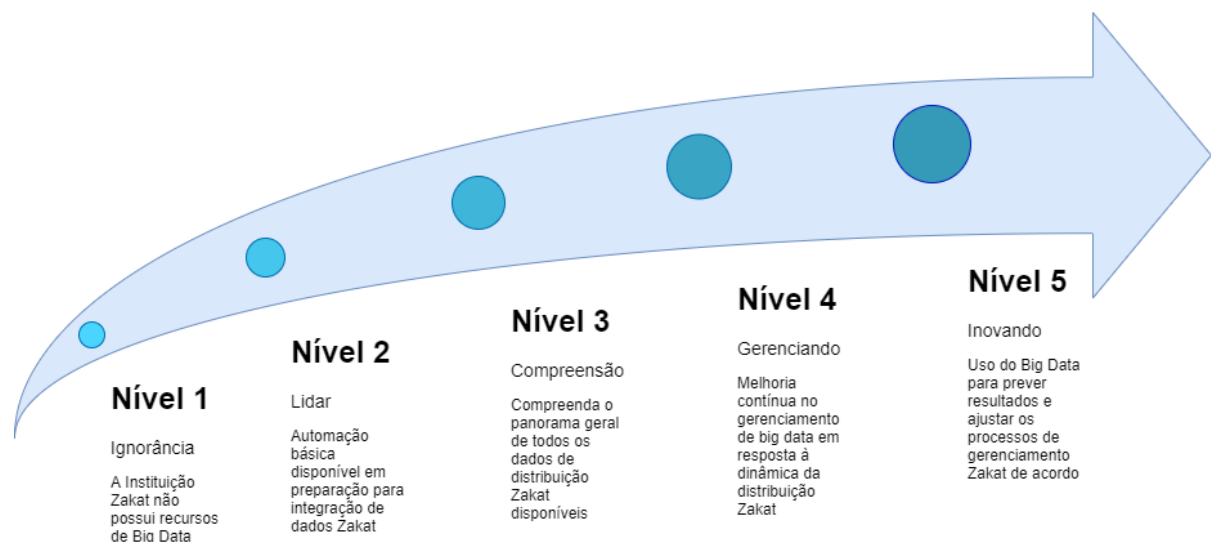
⁴ <https://www.plugar.com.br/data-literacy-voce-conhece-o-conceito-de-alfabetizacao-de-dados/>

3.2.6 Zakat Big Data Maturity Model

Sulaiman, Cob e Ali (2015) propõem um modelo de maturidade para avaliar a prontidão das instituições Zakat nas iniciativas de Big Data. Além disso, o modelo tem como objetivo fornecer às instituições Zakat processos mais bem estruturados de gerenciamento de grandes volumes de dados, a fim de proporcionar transparência dos relatórios e confiança da comunidade islâmica sobre a gestão e distribuição do Zakat⁵.

O modelo é composto por 5 níveis de maturidade: 1 - *Ignorance*, 2 - *Coping*, 3 - *Understanding*, 4 - *Managing* e 5 - *Innovating*. Os autores argumentam, ainda, que para que um nível seja alcançado, é necessário abordar e avaliar todos os obstáculos de cada uma das estratégias dos níveis anteriores a fim de atender a todas as capacidades do processo de acordo com as mudanças entre cada um dos níveis (ver Figura 11).

Figura 11 – *Zakat Big Data Maturity Model*.



Fonte: Extraído de (SULAIMAN; COB; ALI, 2015)

O modelo de Sulaiman, Cob e Ali (2015) fornecerá uma visão geral das iniciativas de big data, para qual estágio de maturidade a instituição precisa chegar e como alcançar tal maturidade. Desta forma, os autores acreditam que os benefícios do destinatário do Zakat aumentarão, bem como o desempenho comercial e a produtividade dessas instituições.

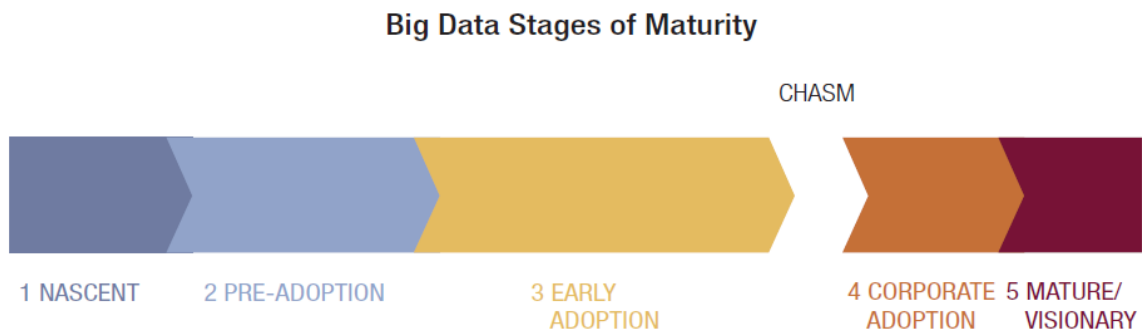
3.2.7 TDWI Big Data Maturity Model

Halper e Krishnan (2013) propõem um modelo com o objetivo de fornecer toda a estrutura necessária para que as organizações consigam ter uma visão de onde estão, onde estiveram e o que ainda precisam realizar no contexto de Big Data. O modelo é dividido em 5 dimensões: *Organization*, *Infrastructure*, *Data Management*, *Analytics* e *Governance*.

⁵ Tributo religioso islâmico. É obrigatório para todos os muçulmanos.

O modelo consiste em 5 níveis de maturidade: 1 - *Nascent*, 2 - *Pre-adoption*, 3 - *Early adoption*, 4 - *Corporate adoption* e 5 - *Mature/Visionary* (ver Figura 12). Onde cada estágio possui uma série de características que o identificam, por exemplo, no estágio 1, as características da organização incluem: gerenciamento de dados, infraestrutura e *analytics*. No estágio 2, além das características presentes no estágio 1, há características a respeito da organização e de governança.

Figura 12 – Níveis de Maturidade do *TDWI Big Data Maturity Model*.



Fonte: Extraído de (HALPER; KRISHNAN, 2013)

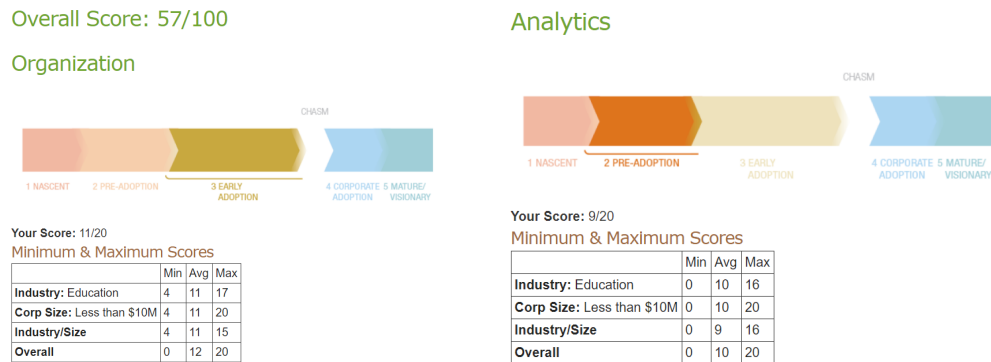
Os autores argumentam ainda que as organizações no estágio 3 permanecem nesse estágio por um tempo maior do que as demais fases. E que uma organização pode ser “madura/visionária” em um aspecto de big data, mas relativamente imatura (estágio 1) em outros aspectos.

Uma avaliação completa com base neste modelo fornecerá à organização uma pontuação em cada dimensão e uma pontuação total, bem como uma recomendação de preenchimento de lacunas, além de fornecer aconselhamento e melhores práticas para chegar ao próximo estágio de maturidade. Ao todo o *survey* conta com 35 questões fornecidas individualmente ou agrupadas em uma matriz. Cada dimensão tem uma pontuação máxima de 20 pontos.

A Figura 13 exemplifica o resultado da avaliação da maturidade de uma organização hipotética onde para a Dimensão *Organization* ela pontuou 11/20 ficando no nível de maturidade “*Early Adoption*” e na Dimensão *Analytics* ela obteve a pontuação 9/20, ficando no nível de maturidade “*Pre-adoption*”

Assim como o método de avaliação do ODMM, em nossa opinião, o método de avaliação do TWDI fornece mais *insights* a seus usuários do que precisa ser melhorado para obtenção de níveis mais altos de maturidade do que a descrição dos níveis de maturidade em si.

Figura 13 – Avaliação da Maturidade *TDWI Assessment*.



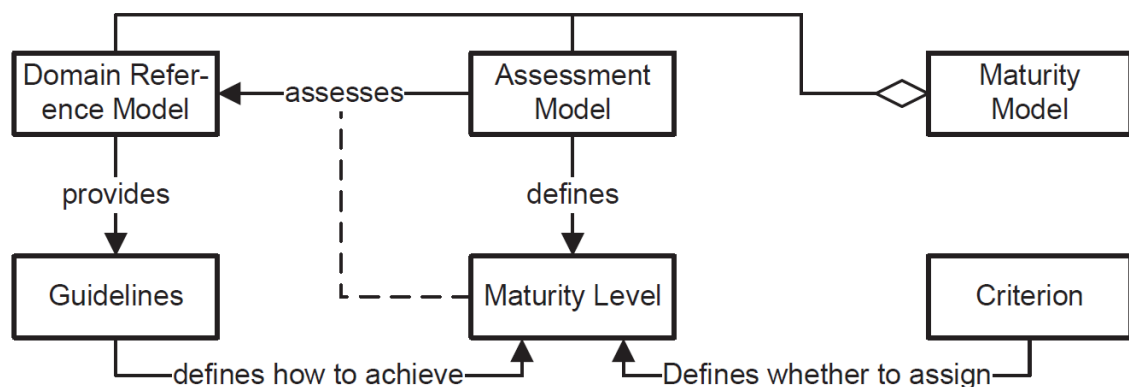
Fonte: Extraído de <<https://tdwi.org>>

3.3 CONSIDERAÇÕES FINAIS

Neste Capítulo foram apresentados alguns modelos de maturidade encontrados durante nossa revisão *ad-hoc* da literatura. Alguns já consolidados e bastante conhecidos como o CMMI, outros de estrutura mais formal e complexa como o modelo multidimensional BPMM, outros puramente descritivo como o DLMM.

Em nossa revisão *ad-hoc* da literatura encontramos dois metamodelos que definem os elementos básicos que estruturam o modelo de referência ou modelo de maturidade de cada um deles (HÜNER; OFNER; OTTO, 2009; NETO et al., 2015). O primeiro metamodelo foi proposto por Hüner, Ofner e Otto (2009) e mostra os componentes e as relações entre cada um deles. O autor ainda declara que o metamodelo não é completo, mas serve como um instrumento para prover termos bem definidos e dependencias para o processo de construção do Modelo de Maturidade para Qualidade de Dados Corporativos (ver Figura 14).

Figura 14 – Componentes de um Modelo de Maturidade.

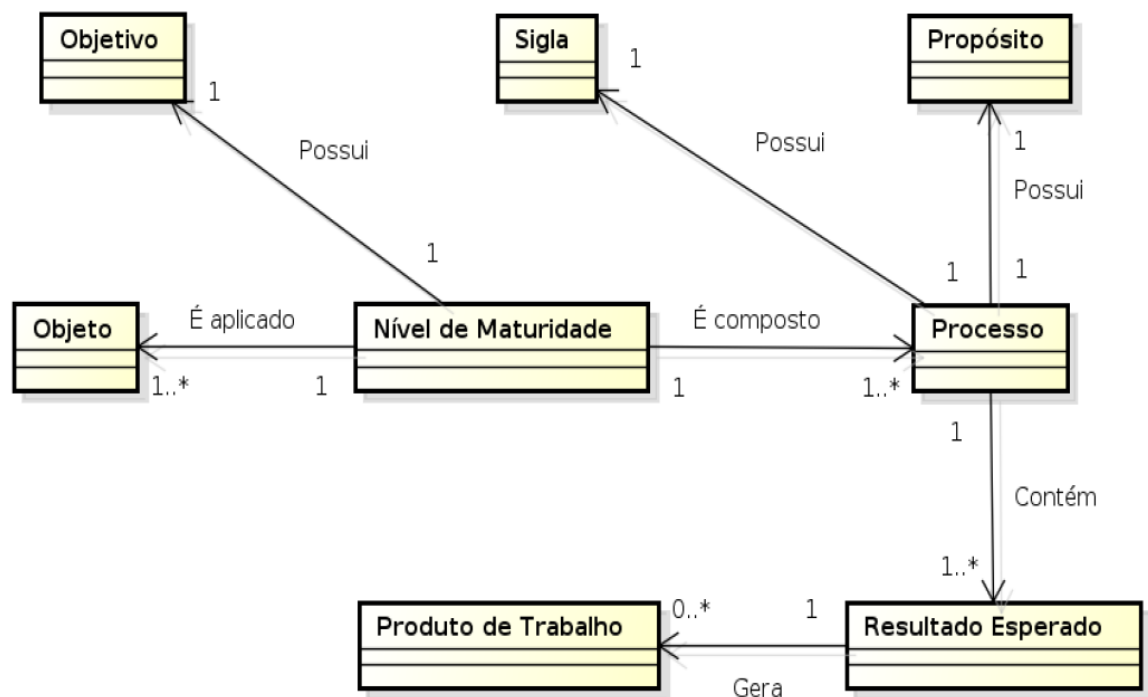


Fonte: Extraído de (HÜNER; OFNER; OTTO, 2009)

Um modelo de maturidade pode ser dividido em duas grandes partes: o modelo de referência e o modelo de avaliação. O modelo de referência provê *guidelines* de como se atingir determinado nível de maturidade. Pode-se dizer que é a estrutura formal que conhecemos em um modelo de maturidade. As descrições das dimensões ou áreas de processo, as práticas, elementos e atividades. Já o modelo de avaliação avalia o desempenho de determinada organização a partir do modelo de referência, definindo em que nível de maturidade tal organização se encontra.

O segundo metamodelo, proposto por Neto et al. (2015), foi criado em forma de modelo conceitual para definir os elementos básicos que estruturam o modelo de referência do *MANGve Maturity Model* (M3) proposto em (NETO, 2015). Neste metamodelo, cada nível de maturidade possui um único objetivo, é composto de um ou mais processos e é aplicado a um ou mais objetos. O objetivo descreve o propósito do nível de maturidade e está relacionado a um único nível de maturidade. O objeto pode ser visto como uma organização inteira ou uma área específica da mesma. O processo pode conter um conjunto de resultados esperados e possui um propósito único. O propósito direciona a organização para a evolução através da descrição de objetivos que norteiam a realização dos processos. E os resultados esperados definem o que pode ser obtido após a execução do processo e podem gerar um conjunto de produtos de trabalho (ver Figura 15).

Figura 15 – Metamodelo do Modelo de Referência *MANGve Maturity Model* (M3).



Fonte: Extraído de (NETO et al., 2015)

Entretanto, como observado dentre os modelos apresentados no capítulo e esses dois

metamodelos, todos compartilham de uma estrutura em comum, como níveis de maturidade, um descritor para cada nível, uma descrição genérica com características de cada nível como um todo, um número de dimensões ou áreas de processos, um número de elementos ou atividades para cada área de processo e uma descrição para cada atividade e/ou como pode ser executada em cada nível de maturidade.

Diante disso, elaboramos um quadro comparativo (ver Quadro 4) entre os modelos de maturidade abordados nesse capítulo, a fim de apresentar as principais semelhanças e diferenças entre eles.

Quadro 4 – Comparação entre os modelos de maturidade encontrados na revisão ad-hoc da literatura

Modelo	Tópico Abordado	Dimensões / Áreas de Processo	Níveis	Natureza	Representação
<i>Capability Maturity Model Integration SEI (2010)</i>	Engenharia de Software	32 áreas de processo	5 níveis de maturidade 4 níveis de capacidade	Descritiva e Prescritiva	Contínua e por Estágios
<i>The Business Process Maturity Model Fisher (2004)</i>	Processos de Negócio	2 grandes dimensões	5 alavancas de mudança 5 estados de maturidade do processo	Descritiva	Por Estágios
<i>Open Data Maturity Model Dodds e Newman (2015)</i>	Dados Abertos	5 temas	5 níveis de maturidade	Descritiva	Contínua e por Estágios
<i>Open Government Maturity Model Lee e Kwak (2012)</i>	Dados Abertos Governamentais	12 temas	5 níveis de maturidade	Descritiva	Por Estágios
<i>Data Literacy Maturity Model Gray, Chambers e Bounegru (2012)</i>	Dados	11 dimensões	4 níveis de maturidade	Descritiva	Por Estágios
<i>Zakat Big Data Maturity Model Sulaiman, Cob e Ali (2015)</i>	Big Data	5 dimensões	5 níveis de maturidade	Descritiva	Por Estágios
<i>TDWI Big Data Maturity Model Halper e Krishnan (2013)</i>	Big Data	5 dimensões	5 níveis de maturidade	Descritiva	Contínua

Fonte: A autora (2019)

4 MODELO DE MATURIDADE PARA DADOS NA WEB

Este capítulo apresenta o modelo de maturidade proposto nesta pesquisa. Para tal, foram estruturadas as seguintes seções: A Seção 4.1 apresenta os principais aspectos do modelo de maturidade proposto abordando, principalmente, sua arquitetura e principais componentes. A Seção 4.2 apresenta os aspectos metodológicos utilizados para o desenvolvimento do modelo, revisão de modelos de maturidade existentes, avaliação e disseminação da pesquisa. A Seção 4.4 apresenta os três níveis de maturidade propostos por este trabalho e a Seção 4.5 apresenta as seis dimensões propostas pelo DWMM, bem como seus elementos em cada nível de maturidade. Por fim, a Seção 4.3 classifica o DWMM e a Seção 4.6 apresenta breves considerações a respeito do que foi apresentado no Capítulo.

4.1 VISÃO GERAL DO DWMM

Nos últimos anos, houve um aumento significativo no volume de dados dispostos na Web. Isso se deve ao crescente interesse em utilizar a Web como canal de compartilhamento e troca de informação. Em sites como DataHub¹, por exemplo, encontramos diversas fontes de dados dos mais variados domínios, muitas delas em formato aberto e/ou conectado. Além disso, a vasta quantidade de dados gerados por meio das redes sociais confirma o potencial da Web como plataforma de compartilhamento e uma grande ferramenta de colaboração entre seus usuários (BOUQUET; STOERMER, 2012).

Apesar deste crescente interesse na publicação de dados na Web e diversas iniciativas propostas para facilitar a publicação e o consumo de dados na Web, ainda encontramos diversos conjuntos de dados difíceis de serem utilizados. De acordo com Santos et al. (2018) esta dificuldade ocorre devido a falhas em aspectos como confiabilidade, reusabilidade, qualidade e interoperabilidade dos conjuntos de dados.

A partir de estudos incluindo, entre outros, um mapeamento sistemático do estado da arte em publicação e consumo de dados na Web (SANTOS et al., 2018), e da análise de soluções para catalogar e publicar dados (OLIVEIRA; OLIVEIRA; LÓSCIO, 2017; SILVA et al., 2015), foi possível observar que o problema de proposição de modelos de maturidade com foco em dados na Web não foi totalmente explorado na literatura. Além disso, mesmo a proposição de metodologias para publicação de dados na Web não é explorada em profundidade, sendo a maioria das alternativas focadas em dados abertos ou dados conectados.

Nesse contexto, o Modelo de Maturidade para Dados na Web (*Data on the Web Maturity Model* - DWMM) foi proposto com o objetivo de avaliar a maturidade de organizações ou indivíduos no processo de publicação de dados na Web. Para atingir tal objetivo, as fa-

¹ <https://datahub.io/>

ses do Modelo de Ciclo de Vida para Dados na Web (DWLM) proposto por (SILVA, 2019) são utilizadas como dimensões do modelo de maturidade proposto. Além disso, para cada uma das etapas do DWLM, ou dimensões do DWMM, foram elencados: (i) os principais desafios enfrentados ao se compartilhar dados na Web, e (ii) as boas práticas que podem ser aplicadas a fim de lidar com tais desafios.

Dessa forma, espera-se que ao avaliar como os desafios dos dados na Web são enfrentados em cada uma das etapas do DWLM seja possível estabelecer o nível de maturidade do processo de publicação propriamente dito. Além disso, conforme discutido em (LÓSCIO; BURLE; CALEGARI, 2016), os desafios dos dados na Web descrevem aspectos fundamentais do compartilhamento de dados na Web e, portanto, devem ser tratados de forma a garantir a publicação de conjuntos de dados que sejam, de fato, úteis para seus potenciais consumidores. Ou seja, espera-se que o tratamento adequado de cada um dos desafios aumente o grau de utilidade dos conjuntos de dados. Neste trabalho, a utilidade de um conjunto de dados é proporcional aos benefícios alcançados ao se aplicar as Boas Práticas para Dados na Web. De acordo com a recomendação DWBP, cada benefício representa uma melhoria em como os conjuntos de dados estão disponíveis na Web.

4.2 METODOLOGIA

Apesar da grande quantidade de modelos de maturidade desenvolvidos tanto por profissionais como por acadêmicos nos últimos anos, não encontramos durante nossa revisão *ad-hoc* da literatura, muitos trabalhos que documentem o processo de desenvolvimento de um modelo de maturidade. Geralmente o desenvolvimento de *frameworks* de maturidade são considerados como *Design Science Research*, uma vez que modelos de maturidade podem ser entendidos como artefatos (BECKER; KNACKSTEDT; PÖPPELBUSS, 2009).

O conceito de *Design Science* foi primeiramente trazido por Herbert Simon, em seu livro intitulado “*The Sciences of the Artificial*”, onde ele apresenta as diferenças entre a ciência natural e a ciência do artificial. Para Simon (1996), a ciência natural seria aquela que se ocupa de descrever e ensinar como os fenômenos naturais funcionam, interagem com o mundo; já “a ciência do artificial se ocupa da concepção de artefatos que realizem objetivos” (SIMON, 1996, p. 114).

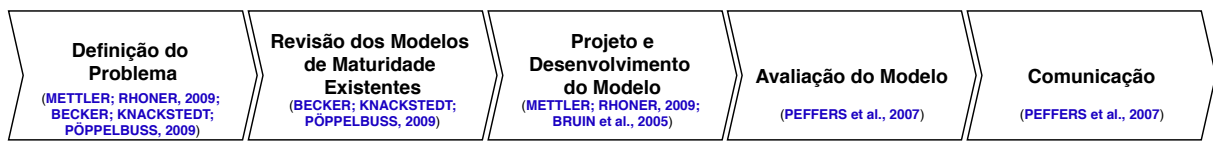
O principal objetivo da *Design Science Research* é desenvolver soluções baseadas em tecnologias produzindo um artefato viável na forma de um construto, um modelo ou instanciações (HEVNER et al., 2004; WENDLER, 2012; BECKER; KNACKSTEDT; PÖPPELBUSS, 2009). Ainda para Becker, Knackstedt e Pöppelbuß (2009), modelos de maturidade podem ser entendidos como artefatos que são utilizados para resolver problemas a partir de medidas de melhoria.

Bruin et al. (2005) propõe um *framework* padrão para o desenvolvimento de modelos de maturidade de qualquer domínio, podendo ser construído de modo descritivo, prescritivo ou comparativo, através de 6 etapas. Becker, Knackstedt e Pöppelbuß (2009)

propõem, com base nas *guidelines* da *Design Science Research* de (HEVNER et al., 2004), um modelo de procedimento genérico e consolidado para o desenvolvimento de modelos de maturidade. Peffers et al. (2007) ainda consolida as propostas de vários autores para a solução de problemas e a construção de artefatos definidas sob o paradigma da *Design Science*.

Tendo em vista que muitas pesquisas necessitam de uma junção de diferentes modelos e métodos para um melhor entendimento do problema, neste trabalho utilizaremos etapas de alguns dos métodos e metodologias já mencionados acima. A Figura 16 descreve as etapas da metodologia de desenvolvimento do DWMM.

Figura 16 – Etapas da Metodologia de Desenvolvimento do DWMM.



Fonte: A autora (2019)

4.2.1 Definição do Problema

Com o intuito de ter uma visão geral do cenário de publicação e consumo de dados, Santos et al. (2018) realizou um mapeamento sistemático para (i) identificar e analisar como os dados têm sido publicados e consumidos na Web, (ii) descobrir os benefícios e limitações da publicação e do consumo de dados na Web, (iii) analisar a evolução da pesquisa sobre publicação e do consumo de dados na Web e (iv) classificar os estudos em categorias relacionadas à sua contribuição. Dentre as áreas que necessitam de mais pesquisa no desenvolvimento e evolução da publicação e consumo de dados na Web, Santos et al. (2018) cita Modelos de Maturidade.

Tendo em vista essa lacuna, identificamos a necessidade de um modelo de maturidade para avaliar as organizações ou indivíduos no processo de publicação de conjuntos de dados na Web. Para isso, será avaliado o processo que foi adotado para publicação de um ou mais conjuntos de dados, ou seja, faz-se necessário analisar todas as etapas do ciclo de vida dos dados na Web, bem como os desafios enfrentados em cada fase do DWLM e como as boas práticas podem ser aplicadas para enfrentar esses desafios e produzir benefícios aos provedores e consumidores de dados.

4.2.2 Revisão dos Modelos de Maturidade Existentes

Na revisão *ad-hoc* da literatura encontramos alguns modelos de maturidade no contexto de dados de uma maneira geral e voltados ao contexto da Web. Alguns dos modelos encontrados já foram descritos no Capítulo 3, dentre os quais podemos destacar dois modelos na área de Dados na Web de uma forma geral, o modelo apresentado por (DODDS;

NEWMAN, 2015) e por (LEE; KWAK, 2012), visto que são modelos de maturidade que tratam a respeito de Dados Abertos e Governo Aberto, respectivamente.

Apesar de existirem modelos que focam em Dados Abertos, como os modelos propostos por Dodds e Newman (2015) e Lee e Kwak (2012), esses modelos não são genéricos o suficiente para o contexto de Dados na Web e em geral, esses modelos focam em aspectos organizacionais, o que não é o foco do modelo proposto por este trabalho.

Embora não tenhamos encontrado modelos de maturidade que avaliem o processo de publicação de dados na Web, a revisão da literatura foi de suma importância para a definição do modelo proposto neste trabalho, uma vez que nos forneceu uma base teórica na área de publicação e consumo de dados e modelos de maturidade.

4.2.3 Projeto e Desenvolvimento do Modelo

A etapa de Projeto e Desenvolvimento do Modelo diz respeito ao processo de construção do modelo de maturidade propriamente dito. As principais decisões a serem tomadas são relativas ao foco do modelo, que pode ser geral ou de um domínio específico.

Becker, Knackstedt e Pöppelbuß (2009) sugere quatro estratégias de desenvolvimento básicas para o Projeto de modelos de maturidade que podem ser: (i) um modelo de maturidade completamente novo, (ii) aprimorar um modelo de maturidade já existente, (iii) combinar vários modelos em um modelo novo e (iv) transferir estruturas e conteúdos de modelos existentes para novos domínios de aplicação. No caso do DWMM, utilizamos alguns conceitos e estruturas já existentes para criação de um modelo de maturidade que atendessem ao processo de publicação de dados na Web de uma maneira geral.

Fraser, Moultrie e Gregory (2002) identifica seis componentes comuns que podem ser encontrados nos modelos de maturidade existentes: (i) um número de níveis (tipicamente entre 3 e 6), (ii) um descritor para cada nível (*e.g.* inicial / repetível / definido / gerenciado / otimizado), (iii) uma descrição genérica ou um resumo das características de cada nível como um todo, (iv) um número de dimensões ou áreas de processo, (v) um número de elementos ou atividades para cada área de processo ou dimensão e (vi) uma descrição para cada atividade que pode ser realizada em cada nível de maturidade. Todos esses componentes serão vistos nas seções 4.4 e 4.5

4.2.4 Avaliação do Modelo

A avaliação de um modelo pode ser realizada por meio de diversas ferramentas, uma delas seria a realização de um grupo focal (LEE; KWAK, 2012; NETO, 2015). Um grupo focal pode ser definido como um grupo de pessoas que se reúnem para avaliar conceitos e/ou evidências de problemas (CAPLAN, 1990).

O grupo focal, como técnica de coleta de dados, já é conhecido e utilizado há várias décadas (BACKES et al., 2011). Uma de suas maiores riquezas é basear-se na tendência humana de formar opiniões e atitudes na interação com outros indivíduos. As pessoas,

em geral, precisam ouvir as opiniões dos outros antes de formar as suas próprias, e consequentemente fundamentam melhor sua posição inicial quando expostas à discussão em grupo (KRUEGER, 2014; MORGAN, 1996).

Nesse contexto, a avaliação do DWMM se dará a partir de um grupo focal a fim de refinar o modelo proposto e assegurar a relevância deste trabalho.

4.2.5 Comunicação

Por fim, a última etapa da metodologia de desenvolvimento do modelo de maturidade diz respeito à divulgação do artefato gerado. No nosso caso, o DWMM será divulgado em meio a comunidade científica por meio de publicações em conferências e periódicos, e por meio desta Dissertação de Mestrado. Além disso, o modelo poderá ser publicado e disponibilizado para organizações que publicam dados na Web. Posteriormente, também será criado um site para disponibilizar a documentação sobre o modelo e servir como fonte de referência, além de ser um canal de comunicação para troca de experiências nos interessados do DWMM.

4.3 CARACTERÍSTICAS E CLASSIFICAÇÃO

Em uma revisão da literatura, Lasrado, Vatrappu e Andersen (2015) identificam que pesquisadores e profissionais acham muito difícil encontrar um modelo de maturidade adequado e pronto para usar que tenha sido validado entre a vasta disponibilidade de literatura. E um dos motivos encontrados é a falta de nomenclatura padrão para descrição do modelo. Dito isto, utilizaremos algumas características mencionadas na literatura por diversos autores a fim de classificar e descrever o DWMM.

4.3.1 Características Gerais

Mettler, Rohner e Winter (2010), em sua proposta de sistema de classificação para modelos de maturidade, definem como atributos gerais do modelo os atributos ou características que são usadas para fornecer rapidamente aos possíveis usuários uma visão geral do modelo de maturidade.

Lasrado, Vatrappu e Andersen (2015) em sua proposta de nomenclatura padrão para descrição do modelo de maturidade, chamam de conceitos básicos alguns dos atributos descritos por Mettler, Rohner e Winter (2010) e inclui mais alguns atributos para compor as características gerais de um modelo de maturidade.

Em nosso trabalho, faremos uma junção dessas duas visões para classificar o DWMM de uma forma geral, e essa classificação pode ser vista no Quadro 5.

Quadro 5 – Características Gerais do DWMM

Atributos	Descrição
Nome	<i>Data on the Web Maturity Model</i>
Acrônimo	<i>DWMM</i>
Foco do Modelo	<i>Domínio Específico</i>
Tópico Abordado	<i>Dados na Web</i>
Origem	<i>Acadêmica</i>
Entidade para Maturação	<i>Processo</i>
Ano de Publicação	<i>2019</i>
Acesso	<i>Livre</i>

Fonte: Adaptado de (LASRADO; VATRAPU; ANDERSEN, 2015; METTLER; ROHNER; WINTER, 2010)

4.3.2 Características de Projeto do Modelo de Maturidade

Para (METTLER; ROHNER; WINTER, 2010) as características de projeto do modelo de maturidade são usadas para descrever a forma e a organização de um modelo de maturidade. Em seu trabalho ele propõe quatro conceitos: (i) atributos de maturidade, (ii) composição, (iii) confiabilidade e (iv) mutabilidade. Lasrado, Vatrappu e Andersen (2015) por sua vez, subdivide as características de design do modelo em três áreas: (i) representação, (ii) construtos e (iii) avaliação. Para esta classificação utilizaremos alguns atributos propostos por Lasrado, Vatrappu e Andersen (2015) (ver Quadro 6).

Quadro 6 – Características de Projeto do DWMM

	Atributos	Descrição
Representação	Composição	<i>Maturity Grids</i>
	Caminho da Maturação	<i>Unidirecional</i>
Construtos	Nível de Abstração	<i>1 - Ad-hoc; 2 - Parcial; 3 - Definido</i>
	Níveis de Maturidade	<i>3</i>
	Representação	<i>Por Estágios</i>
	Natureza do Modelo	<i>Descritivo e Prescritivo</i>
	Dimensões	<i>6</i>
	Elementos	<i>Entre 1 e 6 por Dimensão</i>
Avaliação	Instanciação	<i>Autoavaliação</i>

Fonte: Adaptado de (LASRADO; VATRAPU; ANDERSEN, 2015; METTLER; ROHNER; WINTER, 2010)

4.4 DESCRIÇÃO DOS NÍVEIS DO DWMM

Segundo Fraser, Moultrie e Gregory (2002) e Mettler e Rohner (2009) todos os modelos de maturidade compartilham características comuns, uma delas é o número de níveis de maturidade ou capacidade (tipicamente entre 3 a 6 níveis) além de um descritor para cada nível e isto também foi confirmado durante nossa revisão *ad-hoc* da literatura. Para o contexto do DWMM, definimos neste trabalho três níveis de maturidade para as dimensões e elementos do DWMM, descritos a seguir:

- *Nível 1 - Ad-hoc*

Neste nível, as etapas do ciclo de vida dos dados na Web são desconhecidas, assim como os desafios e as boas práticas para os dados na Web. Desta forma, o processo de publicação de dados é feito de maneira *ad-hoc*.

- *Nível 2 - Parcial*

Neste nível, as etapas do ciclo de vida dos dados na Web já são conhecidas, assim como alguns desafios e algumas boas práticas para dados na Web são adotadas.

- *Nível 3 - Definido*

Neste nível, as etapas do ciclo de vida dos dados na Web já são conhecidas, assim como todos os desafios e todas as boas práticas para os dados na Web são adotadas.

Vale salientar que, por possuir representação por estágios, o Nível 3 compreende tudo que está no Nível 2 e no Nível 1.

4.5 DIMENSÕES DO DWMM

Como mencionado no Capítulo 3, um dos componentes essenciais de um modelo de maturidade são as Dimensões, também conhecidas como Áreas de Processo e Domínios do Modelo, que possuem elementos que, por sua vez, podem ser denominados de Atividades e Atributos (FRASER; MOULTRIE; GREGORY, 2002; CARALLI; KNIGHT; MONTGOMERY, 2012).

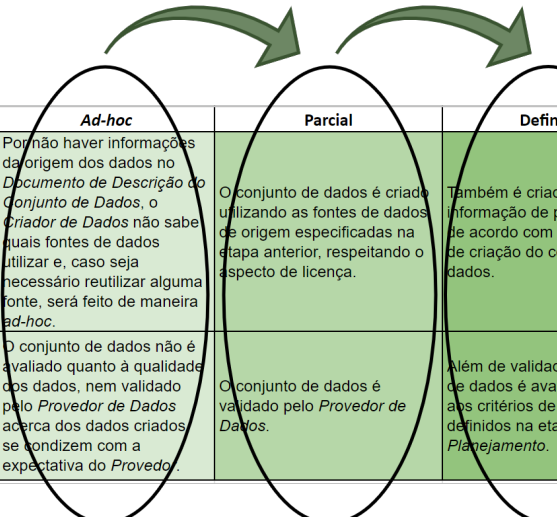
Para Caralli, Knight e Montgomery (2012), as dimensões de um modelo de maturidade (que ele cita como domínios em seu trabalho) são um meio de agrupar atributos semelhantes em uma área de importância para o assunto e a intenção do modelo. Elementos (que ele cita como atributos) por sua vez, representam o conteúdo principal do modelo agrupado por dimensões e níveis. Para Pádua (2003), uma dimensão (que ele cita como área de processo) é um conjunto de práticas relacionadas a uma área que, quando executadas coletivamente, satisfazem a um conjunto de metas consideradas importantes para realização de melhorias significativas.

Neste trabalho, utilizaremos como Dimensões para o DWMM as fases do DWLM proposto por Silva (2019) que engloba seis fases, que vão desde o planejamento até a remoção do acesso aos dados; cinco papéis para os atores que irão, de alguma forma, interagir com os dados ao longo do ciclo de vida; e dezessete atividades distribuídas entre as fases do ciclo.

Através de um conjunto de casos de uso² selecionados pelo grupo de trabalho do W3C foi possível identificar os principais desafios enfrentados por provedores e consumidores de dados ao longo do ciclo de vida dos dados na Web (LÓSCIO; GUIMARÃES; CALEGARI, 2016). Com isso, verificamos a necessidade de analisar como os desafios dos dados na Web estão sendo enfrentados ao longo do processo de publicação dos dados na Web, para tal identificamos os desafios enfrentados em cada fase do ciclo de vida dos dados a partir das atividades definidas por (SILVA, 2019) em cada etapa do DWLM e utilizamos como os elementos de cada dimensão no DWMM (ver Figura 18).

É importante ressaltar que o DWMM possui uma representação por estágios, ou seja, não é possível avaliar a maturidade de um único elemento (processo) sozinho. Desta forma, para cada dimensão será avaliado o conjunto de desafios distribuídos em cada nível de maturidade e como as boas práticas são adotadas para enfrentar tais desafios. Para um melhor entendimento, ver Figura 17.

Figura 17 – Representação por Estágios do DWMM.



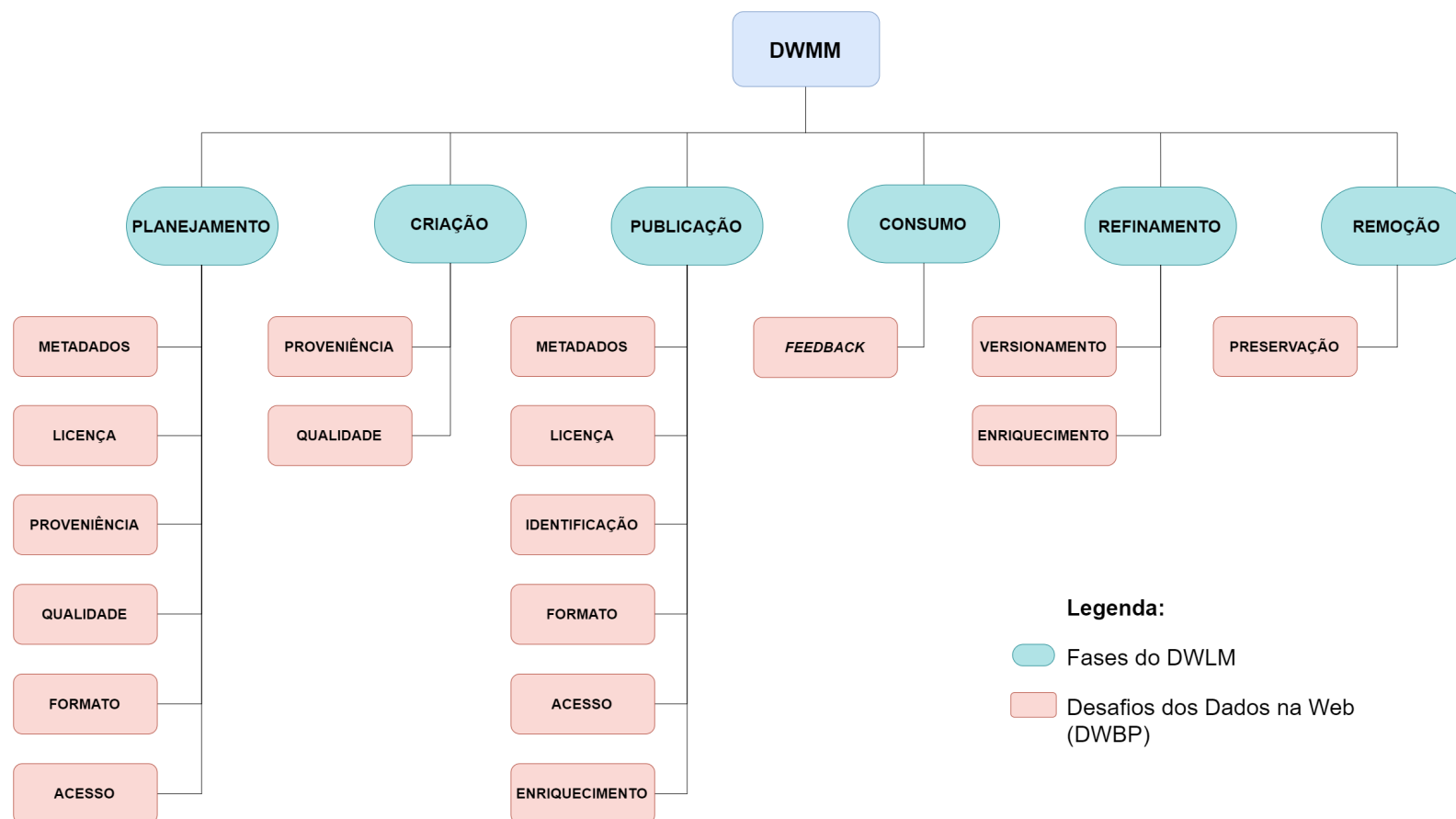
Dimensão ↓	Elementos	Ad-hoc	Parcial	Definido
Criação	Proveniência	Por não haver informações da origem dos dados no Documento de Descrição do Conjunto de Dados, o Criador de Dados não sabe quais fontes de dados utilizar e, caso seja necessário reutilizar alguma fonte, será feito de maneira ad-hoc.	O conjunto de dados é criado utilizando as fontes de dados de origem especificadas na etapa anterior, respeitando o aspecto de licença.	Também é criada a informação de proveniência de acordo com o processo de criação do conjunto de dados.
	Qualidade	O conjunto de dados não é avaliado quanto à qualidade dos dados, nem validado pelo Provedor de Dados acerca dos dados criados se condizem com a expectativa do Provedor.	O conjunto de dados é validado pelo Provedor de Dados.	Além de validado, o conjunto de dados é avaliado quanto aos critérios de qualidade definidos na etapa de Planejamento.

Fonte: A autora (2019)

Nas próximas subseções, cada uma das dimensões do DWMM será explicada juntamente com a descrição correspondente dos níveis de maturidade, elementos de cada dimensão (i.e., desafios enfrentados em cada fase) e as boas práticas associadas a cada dimensão.

² <https://www.w3.org/TR/dwbp-ucr>

Figura 18 – Dimensões e Elementos do DWMM.



A autora (2019)

4.5.1 Planejamento

Segundo (SILVA, 2019), a fase de planejamento é de suma importância, pois é onde ocorre a descrição do conjunto de dados, bem como a definição dos metadados que serão disponibilizados juntamente com o conjunto. Além disso, informações sobre a proveniência, licença e volume do conjunto de dados também são obtidas nesta fase.

Considerando os objetivos do DWMM, os desafios que surgem durante a etapa de Planejamento do DWLM foram identificados a partir das 3 atividades (*Especificar fontes de dados*, *Descrever o conjunto de dados* e *Estabelecer solução para publicação do conjunto de dados*) propostas por (SILVA, 2019) para esta etapa e são descritos no Quadro 7

Quadro 7 – Desafios da Dimensão Planejamento

Dimensão DWMM: Planejamento	
Desafio DWBP	Desafio DWMM
Metadados	Realizar um planejamento de quais metadados serão disponibilizados aos consumidores e documentar isso no Documento de Descrição do Conjunto de Dados.
Licença	Analisar as licenças de dados disponíveis e escolher a licença a ser usada para o conjunto de dados; se o mesmo poderá ser reutilizado, modificado e/ou compartilhado.
Proveniência	Especificar as fontes de dados que poderão ser usadas para a coleta dos dados que serão publicados e descrever o processo de coleta e extração dos dados.
Formato	Especificar quais formatos de dados serão utilizados para a criação do conjunto de dados; de preferência, estabelecer múltiplos formatos.
Acesso	Estabelecer a solução para publicação do conjunto de dados e especificá-la no Documento de Descrição do Conjunto de Dados e definir a frequência de atualização do conjunto de dados.

Fonte: A autora (2019)

No Quadro 8, podemos ver as Saídas esperados para essa Dimensão, quais Boas Práticas são/devem ser aplicadas para cada nível e os benefícios gerados por elas de forma distinta.

Ainda, durante o processo de publicação de dados na Web surgem alguns desafios que precisam ser enfrentados pelos atores envolvidos no processo. Tais desafios podem ser superados por meio da utilização das boas práticas para dados na Web. Lóscio, Guimarães e Calegari (2016) afirmam que a utilização das boas práticas pode ajudar os provedores e consumidores a alcançarem uma série de benefícios que se tornam mais fortes na medida em que a adoção das boas práticas aumentam.

Desta forma, embora o conjunto de dados ainda não tenha sido propriamente criado, as ações realizadas nessa fase, adotando as boas práticas, resultarão em alguns benefícios que caracterizarão a utilidade do conjunto de dados. Os benefícios são listados distintamente no Quadro 8.

Quadro 8 – Dimensão Planejamento do DWMM

PLANEJAMENTO			
Nível	Saídas	Boas Práticas Aplicadas	Benefícios
<i>Ad-hoc</i>	Não há planejamento definido; Pouca ou nenhuma informação sobre metadados, licença e proveniência dos dados estão disponíveis no <i>Documento de Descrição do Conjunto de Dados</i>; Solução primária para disponibilização de dados é definida;	Nenhuma	Nenhum
<i>Parcial</i>	O Documento de Descrição do Conjunto de Dados contém informações sobre as fontes de dados de origem que serão utilizadas para criação do conjunto de dados; O Documento de Descrição do Conjunto de Dados contém informações sobre metadados descritivos, informações de licença, os formatos e distribuições dos dados que serão disponibilizados; A solução definida para disponibilização e acesso aos dados é básica, intermediária ou por meio de uma API;	BP2 - Fornecer Metadados Descritivos BP4 - Fornecer Informações sobre a Licença dos Dados BP5 - Fornecer Informações de Proveniência dos Dados BP6 - Fornecer Informação de Qualidade de Dados BP35 - Citar a Publicação Original do Conjunto de Dados	Reuso Compreensão Descoberta Confiabilidade
<i>Definido</i>	A etapa de planejamento é bem definida; O Documento de Descrição do Conjunto de Dados contém informações sobre metadados descritivos e estruturais, licença, formatos e distribuições dos dados que serão disponibilizados e os critérios de qualidade que serão utilizados na etapa de Criação; É definida uma solução avançada para disponibilização de dados;	BP1 - Fornecer Metadados BP2 - Fornecer Metadados Descritivos BP3 - Fornecer Metadados Estruturais BP4 - Fornecer Informações sobre a Licença dos Dados BP5 - Fornecer Informações de Proveniência dos Dados BP6 - Fornecer Informação de Qualidade de Dados BP35 - Citar a Publicação Original do Conjunto de Dados	Reuso Compreensão Descoberta Confiabilidade Processabilidade

Fonte: A autora (2019)

Levando em consideração os níveis de maturidade e os aspectos já mencionados a respeito dos elementos da dimensão Planejamento, uma visão do DWMM para esta etapa pode ser vista no Quadro 9.

4.5.2 Criação

Esta é a fase em que ocorre de fato a criação dos conjuntos de dados, por meio da extração dos mesmos nas fontes de origem. Silva (2019) cita três cenários em que a extração dos dados acontece e que precisam passar por processos diferentes. No primeiro cenário, os dados não precisam passar por um processo de transformação, integração ou limpeza, dessa forma acontecerá apenas um processo de carga para o armazenamento do conjunto de dados. No segundo cenário, os dados são provenientes de *APIs* em tempo real, deste modo a criação do conjunto de dados consiste em disponibilizar um *link* para que os dados possam ser acessados via Web. No terceiro cenário os dados necessitam passar por processos de transformação, integração ou limpeza, neste caso faz-se necessário a realização de um processo de ETL.

Além da extração dos dados das fontes de origem, na fase de Criação o conjunto de dados passa por um processo de avaliação quanto à sua qualidade através das métricas que o *Criador de Conjunto de Dados* julgar importantes para o contexto do conjunto de dados.

Considerando os objetivos do DWMM, os desafios que surgem durante a etapa de Criação do DWLM foram identificados a partir de 3 atividades (*Criar o conjunto de dados*, *Avaliar qualidade* e *Validar o conjunto de dados recém criado*) propostas por (SILVA, 2019) para esta etapa e são descritos no Quadro 10.

Quadro 10 – Desafios da Dimensão Criação

Dimensão DWMM: Criação	
Desafio DWBP	Desafio DWMM
Proveniência	Criar o conjunto de dados utilizando as fontes de dados que foram especificadas na etapa anterior e definir a estratégia de extração dos dados.
Qualidade	Avaliar o conjunto de dados quanto à sua qualidade com relação aos critérios e métricas definidos na etapa de Planejamento.

Fonte: A autora (2019)

As Saídas esperados para a Dimensão Criação, as boas práticas e os benefícios gerados por elas são descritos, nos três níveis de maturidade propostos pelo DWMM, no Quadro 11.

Levando em consideração os níveis de maturidade e os aspectos já mencionados a respeito dos elementos da dimensão Criação, uma visão do DWMM para esta etapa pode ser vista no Quadro 12.

Quadro 9 – Elementos da Dimensão Planejamento

Dimensão Planejamento			
Elementos	<i>Ad-hoc</i>	Parcial	Definido
Metadados	Não há metadados no <i>Documento de Descrição do Conjunto de Dados</i> .	O conjunto de dados contém apenas metadados descritivos, como título, descrição do conjunto de dados e data de publicação.	O conjunto de dados também contém metadados estruturais e possui informações sobre palavras-chaves que o descrevem, entidade responsável por disponibilizar o conjunto, contato, sua cobertura espacial, período temporal que os dados abrangem, data da última modificação, tema/categoria e frequência de atualização, bem como a estrutura do conjunto de dados para que os consumidores possam entender os significados dos dados.
Licença	Não há informações quanto à Licença dos dados no <i>Documento de Descrição do Conjunto de Dados</i> .	O <i>Documento de Descrição do Conjunto de Dados</i> contém informações de licença.	O <i>Documento de Descrição do Conjunto de Dados</i> contém informações de licença.
Proveniência	Não há informações da origem dos dados no <i>Documento de Descrição do Conjunto de Dados</i> .	Há informações do processo de derivação dos dados, sua origem e seu histórico.	Há informações do processo de derivação dos dados, sua origem e seu histórico.
Qualidade	Não há critérios de qualidade definidos para avaliação do conjunto de dados.	Não há critérios de qualidade definidos para avaliação do conjunto de dados.	O <i>Criador de Conjunto de Dados</i> define quais critérios de qualidade serão utilizados para avaliar o conjunto de dados na fase de Criação.
Formato	Não há informações de especificação dos formatos de dados a serem disponibilizados.	O <i>Documento de Descrição do Conjunto de Dados</i> contém informações sobre os formatos de distribuição para especificação de como os dados serão disponibilizados.	O <i>Documento de Descrição do Conjunto de Dados</i> contém informações sobre os formatos de distribuição para especificação de como os dados serão disponibilizados.
Acesso	O <i>Provedor de Dados</i> estabelece o uso de uma solução primária para disponibilização dos dados na Web (e.g. Uma página HTML) com no mínimo a opção de <i>download</i> do conjunto de dados e de seus subconjuntos.	O <i>Provedor de Dados</i> estabelece o uso de uma solução básica ou intermediária para disponibilização do conjunto de dados onde estão descritos os metadados do conjunto de dados em formato legível por máquina, e/ou a disponibilização de uma API para que os dados possam ser acessados por meio dela.	O <i>Provedor de Dados</i> estabelece o uso de uma solução avançada para disponibilização dos dados na Web onde, além de estender tudo o que a solução intermediária oferece, a mesma possui pesquisas por meio de filtros e buscas por palavras-chaves, pré-visualizações das distribuições do conjunto de dados, ambiente de coleta de <i>feedback</i> dos usuários, suporte ao versionamento do conjunto de dados, dentre outras.

Fonte: A autora (2019)

Quadro 11 – Dimensão Criação do DWMM

CRIAÇÃO			
Nível	Saídas	Boas Práticas Aplicadas	Benefícios
Ad-hoc	O conjunto é criado de maneira <i>ad-hoc</i> ;	Nenhuma	Nenhum
Parcial	O conjunto é criado utilizando as fontes de dados de origem respeitando os aspectos de licença; A informação de proveniência das fontes de origem é registrada; O conjunto é validado pelo Provedor de Dados;	BP5 - Fornecer Informações de Proveniência dos Dados BP34 - Obedecer os Termos de Licença	Reuso Compreensão Confiabilidade
Definido	O conjunto é criado utilizando as fontes de dados de origem respeitando os aspectos de licença; A informação de proveniência sobre o processo de criação do conjunto de dados é registrada; O conjunto é validado pelo <i>Provedor de Dados</i> ; O conjunto é avaliado quanto à qualidade dos dados;	BP5 - Fornecer Informações de Proveniência dos Dados BP6 - Fornecer Informações de Qualidade de Dados BP34 - Obedecer os Termos de Licença	Reuso Compreensão Confiabilidade

Fonte: A autora (2019)

Quadro 12 – Elementos da Dimensão Criação

Criação			
Elementos	<i>Ad-hoc</i>	Parcial	Definido
Proveniência	Por não haver informações da origem dos dados no <i>Documento de Descrição do Conjunto de Dados</i> , o <i>Criador de Dados</i> não sabe quais fontes de dados utilizar e, caso seja necessário reutilizar alguma fonte, será feito de maneira <i>ad-hoc</i> .	O conjunto de dados é criado utilizando as fontes de dados de origem especificadas na etapa anterior, respeitando o aspecto de licença.	Também é criada a informação de proveniência de acordo com o processo de criação do conjunto de dados.
Qualidade	O conjunto de dados não é avaliado quanto à qualidade dos dados, nem validado pelo <i>Provedor de Dados</i> quanto acerca dos dados criados, se condizem com a expectativa do <i>Provedor</i> .	O conjunto de dados é validado pelo <i>Provedor de Dados</i> .	Além de validado, o conjunto de dados é avaliado quanto aos critérios de qualidade definidos na etapa de <i>Planejamento</i> .

Fonte: A autora (2019)

4.5.3 Publicação

Esta é a fase em que os dados serão disponibilizados na Web de acordo com a solução escolhida para sua publicação. As informações descritas no *Documento de Descrição do Conjunto de Dados* serão utilizadas nesta fase para a publicação do conjunto de dados e seus subconjuntos, seus metadados e o vocabulário.

Considerando os objetivos do DWMM, os desafios que surgem durante a etapa de Criação do DWLM foram identificados a partir de 3 atividades (*Publicar o conjunto de dados de acordo com a solução escolhida*, *Tornar o conjunto de dados acessível na Web* e *Fornecer alternativas de uso*) propostas por (SILVA, 2019) para esta etapa e são descritos no Quadro 13.

Quadro 13 – Desafios da Dimensão Publicação

Dimensão DWMM: Publicação	
Desafio DWBP	Desafio DWMM
Metadados	Publicar os metadados descritivos e estruturais documentos no Documento de Descrição do Conjunto de Dados, em formato que seja legível por humanos e por máquinas.
Licença	Disponibilizar ao consumidor dos dados a informação sobre o licenciamento do conjunto de dados, descrevendo quais permissões de acesso e manipulação de dados está disponível ao consumidor; disponibilizando de forma que seja legível por humanos e detectada automaticamente por máquinas.
Formato	Publicar o conjunto de dados em múltiplos formatos que possibilitem sua fácil manipulação e interpretação por humanos e por máquinas.
Identificação	Publicar o conjunto de dados com padrões URLs / URIs
Acesso	Possibilitar o download do conjunto de dados por inteiro em uma única requisição, bem como disponibilizar subconjuntos de dados se o mesmo possuir uma vasta quantidade de dados; o conjunto de dados deverá estar disponível na Web em tempo real ou o mais próximo disso; o conjunto de dados deverá estar atualizado e com informações descritivas sobre os dados que não estão mais disponíveis e como/por quem esse dado pode ser acessado. Se houver recursos disponíveis para tal, disponibilizar os dados por meio de uma API utilizando a infraestrutura/tecnologias da Web.
Enriquecimento	Possibilitar o enriquecimento dos dados apresentando-os de forma complementar e informativa por meio de visualizações, tabelas e exemplos de uso do conjunto de dados a fim de que os consumidores tenham uma visão geral de como os dados podem ser utilizados.

Fonte: A autora (2019)

As Saídas esperados para a Dimensão Publicação, as boas práticas e os benefícios gerados por elas podem são descritos, nos três níveis de maturidade propostos pelo DWMM, no Quadro 14.

Levando em consideração os níveis de maturidade e os aspectos já mencionados a respeito dos elementos da Dimensão Publicação, uma visão do DWMM para esta etapa pode ser vista no Quadro 15.

Quadro 14 – Dimensão Publicação do DWMM

PUBLICAÇÃO			
Nível	Saídas	Boas Práticas Aplicadas	Benefícios
<i>Ad-hoc</i>	O conjunto é publicado em um único formato; É possível realizar o <i>download</i> do conjunto de dados e seus subconjuntos; O conjunto de dados é publicado apenas em um formato <i>human-readable</i>;	BP17 - Fornecer Bulk Download	Reuso Acesso
<i>Parcial</i>	O conjunto de dados é publicado em múltiplos formatos, incluindo pelo menos um formato <i>machine-readable</i>; O conjunto de dados é publicado com uma licença que permite a reutilização dos dados; É possível realizar o <i>download</i> do conjunto de dados e seus subconjuntos; O conjunto de dados é publicado por meio da disponibilização de uma API para acesso aos dados ou por meio de uma solução básica ou intermediária;	BP2 - Fornecer metadados descritivos BP9 - Usar URIs persistentes como identificadores de Conjuntos de Dados BP12 - Usar formatos de dados padronizados legíveis por máquinas BP13 - Usar Representações de dados que sejam independentes de localidade (Locale Neutral) BP14 - Fornecer dados em vários formatos BP17 - Fornecer Bulk Download BP18 - Fornecer subconjuntos para conjuntos de dados grandes BP20 - Fornecer acesso em tempo real BP23 - Tornar os dados disponíveis por meio de uma API BP34 - Obedecer os termos de Licença	Reuso Compreensão Confiabilidade
<i>Definido</i>	O conjunto de dados é publicado em múltiplos formatos e possui uma distribuição em RDF; Metadados descritivos e estruturais são publicados com o conjunto de dados; O conjunto de dados é publicado por meio da disponibilização de uma API ou de uma solução avançada; O conjunto de dados é enriquecido por meio da disponibilização de visualizações complementares; Há URIs como identificadores dentro do conjunto de dados e para todas as versões existentes do conjunto de dados; É possível descobrir novos recursos por meio da utilização de URIs;	BP1 - Fornecer metadados BP2 - Fornecer metadados descritivos BP3 - Fornecer metadados estruturais BP9 - Usar URIs persistentes como identificadores de Conjuntos de Dados BP10 - Usar URIs persistentes como identificadores dentro de conjuntos de dados BP11 - Atribuir URIs para as versões dos conjuntos de dados e séries BP12 - Usar formatos de dados padronizados legíveis por máquinas BP13 - Usar Representações de dados que sejam independentes de localidade (Locale Neutral) BP14 - Fornecer dados em vários formatos BP17 - Fornecer Bulk Download BP18 - Fornecer subconjuntos para conjuntos de dados grandes BP20 - Fornecer acesso em tempo real BP23 - Tornar os dados disponíveis por meio de uma API BP24 - Usar padrões Web como base para construção de APIs BP32 - Fornecer visualizações complementares BP34 - Obedecer os termos de Licença	Reuso Compreensão Descoberta Conectividade Interoperabilidade Processabilidade Acesso Confiabilidade

Quadro 15 – Elementos da Dimensão Publicação

Publicação			
Elementos	<i>Ad-hoc</i>	Parcial	Definido
Metadados	Não há metadados publicados para o conjunto de dados.	São publicados metadados descritivos.	Também são publicados metadados estruturais legíveis por humanos e máquinas.
Licença	O conjunto de dados não possui uma licença associada.	O conjunto de dados é publicado com uma licença que permite a reutilização e compartilhamento dos mesmos, porém é apenas compreendida por humanos.	Além de ser compreendida por humanos, as máquinas conseguem detectar automaticamente as informações sobre a licença dos dados.
Identificação	Há um meio genérico de identificação dos conjuntos de dados por meio de URLs.	Há URIs como nomes de recursos e identificadores do conjunto de dados.	Há também <i>links</i> para outros URIs possibilitando a descoberta de novos recursos, e também são disponibilizadas URIs para versões de conjuntos de dados.
Formato	O conjunto de dados é publicado em um único formato.	O conjunto de dados é publicado em múltiplos formatos.	O conjunto de dados publicado também possui uma distribuição RDF, além de ser conectado com outras bases de dados.
Acesso	O conjunto de dados é publicado por meio de uma solução primitiva.	O conjunto de dados é publicado por meio de uma solução básica ou intermediária.	O conjunto de dados é publicado por meio de uma solução avançada.
Enriquecimento	Não há possibilidade de enriquecimento dos dados por meio de visualizações e exemplos de uso.	Não há possibilidade de enriquecimento dos dados por meio de visualizações e exemplos de uso do conjunto de dados.	É possível enriquecer o conjunto de dados disponibilizando exemplos de uso ao consumidor, e visualizações que possibilitem a fácil compreensão do significado dos dados.

Fonte: A autora (2019)

4.5.4 Consumo

Nesta fase os dados estão disponíveis aos consumidores seja através de APIs ou soluções que permitam a criação de visualizações, como gráficos, tabelas e demais análises de dados possíveis.

A fase de Consumo está diretamente relacionada ao consumidor de dados, porém, neste trabalho, avaliaremos a fase de Consumo do ponto de vista dos atores que exercem alguma ação relacionada à publicação, visto que o nosso objetivo é avaliar os conjuntos

de dados gerados e o processo de publicação de dados. Nesse contexto, o desafio que se relaciona com a fase de Consumo é o de *Feedback* (ver Quadro 16).

Quadro 16 – Desafios da Dimensão Consumo

Dimensão DWMM: Consumo	
Desafio DWBP	Desafio DWMM
<i>Feedback</i>	Prover uma solução de coleta e disponibilização de <i>feedback</i> sobre o conjunto de dados, tornando o <i>feedback</i> acessível a todos os seus consumidores

Fonte: A autora (2019)

As Saídas esperados para a Dimensão Consumo, as boas práticas e os benefícios gerados por elas são descritos, nos três níveis de maturidade propostos pelo DWMM, no Quadro 17.

Quadro 17 – Dimensão Consumo do DWMM

CONSUMO			
Nível	Saídas	Boas Práticas Aplicadas	Benefícios
<i>Ad-hoc</i>	Não há canal de comunicação entre provedores e consumidores de dados;	Nenhuma	Nenhum
<i>Parcial</i>	Há um meio para coleta de <i>feedback</i> ;	BP29 - Coletar <i>feedback</i> dos consumidores de dados	Reuso Compreensão Confiabilidade
<i>Definido</i>	Há um meio para coleta de <i>feedback</i> e o <i>feedback</i> coletado está disponível na Web para todos os consumidores dos dados;	BP29 - Coletar <i>feedback</i> dos consumidores de dados BP30 - Compartilhar o <i>feedback</i> disponível	Reuso Compreensão Confiabilidade

Fonte: A autora (2019)

Uma visão do DWMM para a fase de Consumo pode ser vista no Quadro 18.

Quadro 18 – Elementos da Dimensão Consumo

Consumo			
Elementos	<i>Ad-hoc</i>	Parcial	Definido
<i>Feedback</i>	Não há um meio onde o consumidor possa deixar seu <i>feedback</i> sobre o conjunto de dados publicado.	Há um local para coleta de <i>feedback</i> .	Os <i>feedback</i> coletados estão acessíveis a qualquer consumidor na Web.

Fonte: A autora (2019)

4.5.5 Refinamento

Esta fase compreende as atividades de atualização que o conjunto de dados pode sofrer ao longo de seu ciclo de vida. Esta fase é de suma importância para garantir a atualização e correção do conjunto de dados para aqueles que irão consumir. Este refinamento pode ocorrer seja a partir do *Feedback* coletado na fase de Consumo, por atualizações programadas ou atualizações não programadas.

Em geral, qualquer atualização que um conjunto de dados venha a passar é considerado um refinamento dos dados. Uma atualização é dita programada quando ocorre em consequência da frequência de atualização do conjunto de dados, ou seja, são atualizações realizadas para garantir a consistência do conjunto de dados com sua fonte de origem e ocorrem automaticamente de acordo com uma frequência pré-estabelecida. As atualizações não programadas também são realizadas para garantir a consistência com a fonte de dados de origem, mas não são realizadas em períodos pré-determinados.

Santos (2018) propõe em seu trabalho alguns procedimentos de limpeza e enriquecimento de dados. No processo de limpeza, são definidas algumas operações de busca, correções de erros e remoção de duplicatas, enquanto no processo de enriquecimento são definidas operações de adição de dados e novos valores. Esses procedimentos são listados a seguir:

- Procedimentos de Limpeza
 1. Correção de valores falsos;
 2. Correção de ortografia;
 3. Correção de valores ocultos;
 4. Correção de valores abreviados;
 5. Correção de erros referenciais;
 6. Correção de valores desviados;
 7. Remoção de registros duplicados.

- Procedimentos de Enriquecimento
 1. Adição de dados em atributos com valores vazios;
 2. Adição de atributo;
 3. Adição de registros;
 4. Adição de metadados;
 5. Atualização de metadados;
 6. Anotação semântica de um valor;
 7. Anotação semântica de um metadado.

Santos (2018) ainda recomenda a criação de um documento de *log* que especifique as alterações realizadas durante os procedimentos de correção e/ou enriquecimento do conjunto de dados para que o *Provedor de Dados* possa validar.

Vale ressaltar que uma atividade de refinamento resulta em uma nova versão do conjunto de dados, ou seja, cada procedimento de limpeza e enriquecimento que o *Criador de Conjunto de Dados* executar, seja por meio de uma atualização programada ou não, deverá gerar uma nova versão do conjunto de dados para que o consumidor saiba que está consumindo a versão mais recente daquele conjunto de dados, e a versão anterior ao refinamento deve estar disponível para os consumidores dos dados a qualquer momento.

Considerando os objetivos do DWMM, os desafios que surgem durante a etapa de Refinamento do DWLM foram identificados a partir das 3 atividades (*Corrigir e enriquecer o conjunto de dados*, *Validar o conjunto de dados refinado* e *Versionar o conjunto de dados*) propostas por (SILVA, 2019) para esta etapa e são descritos no Quadro 19.

Quadro 19 – Desafios da Dimensão Refinamento

Dimensão DWMM: Refinamento	
Desafio DWBP	Desafio DWMM
Versionamento	Disponibilizar histórico de atualizações e refinamento ao longo do tempo em que o conjunto de dados foi publicado e permanece disponível para os consumidores na Web e as versões anteriores do conjunto de dados deverão permanecer disponíveis para o consumidor a qualquer momento.
Enriquecimento	Realizar procedimentos de limpeza e enriquecimento a partir do feedback coletado na etapa anterior ou por atualizações pré-estabelecidas no Documento de Descrição do Conjunto de Dados.

Fonte: A autora (2019)

As Saídas esperados para a Dimensão Refinamento, as boas práticas e os benefícios gerados por elas podem são descritos, nos três níveis de maturidade propostos pelo DWMM, no Quadro 20.

Quadro 20 – Dimensão Refinamento do DWMM

REFINAMENTO			
Nível	Saídas	Boas Práticas Aplicadas	Benefícios
<i>Ad-hoc</i>	Não há iniciativas de limpeza e enriquecimento dos dados; Não há histórico de versões disponíveis;	Nenhuma	Nenhum
<i>Parcial</i>	Há iniciativas de limpeza e enriquecimento dos dados por parte de atualizações programadas ou não programadas; Há um histórico descritivo sobre as versões do conjunto de dados;	BP7 - Fornecer indicador de versão BP31 - Enriquecer dados por meio da geração de novos dados	Reuso Compreensão Confiabilidade Processabilidade
<i>Definido</i>	Há iniciativas de limpeza e enriquecimento dos dados por parte de atualizações programadas ou não programadas; É gerado um documento de log com todo o processo de limpeza e refinamento; Todas as versões do conjunto de dados estão disponíveis na Web ao longo do ciclo de vida dos dados;	BP7 - Fornecer indicador de versão BP8 - Fornecer histórico de versões BP31 - Enriquecer dados por meio da geração de novos dados	Reuso Compreensão Confiabilidade Processabilidade

Fonte: A autora (2019)

Levando em consideração os níveis de maturidade e os aspectos já mencionados a respeito dos elementos da dimensão Refinamento, uma visão do DWMM para esta etapa pode ser vista no Quadro 21.

Quadro 21 – Elementos da Dimensão Refinamento

Refinamento			
Elementos	<i>Ad-hoc</i>	Parcial	Definido
Versionamento	Não há um histórico de versões do conjunto de dados.	Há um histórico descritivo das versões do conjunto de dados <i>human and machine readable</i> .	Há um histórico de versões e as versões anteriores atualizadas / refinadas estão disponíveis na Web para os consumidores a qualquer momento.
Enriquecimento	Não existem iniciativas de enriquecimento dos dados.	Existem iniciativas de enriquecimento dos dados através da correção por parte do próprio criador ou do <i>feedback</i> coletado na fase anterior, bem como procedimentos de limpeza e enriquecimento de dados.	É criado um documento de log com a especificação do que foi alterado durante os procedimentos de limpeza e enriquecimento dos dados.

Fonte: A autora (2019)

4.5.6 Remoção

A fase de Remoção finaliza o ciclo de vida dos dados. Ela corresponde ao processo de preservação adequada e a retirada de acesso ao conjunto de dados.

Para esta fase (SILVA, 2019) define apenas uma atividade (*Remover acesso ao conjunto de dados*) onde identificamos o desafio de Preservação (ver Quadro 22).

Quadro 22 – Desafios da Dimensão Remoção

Dimensão DWMM: Remoção	
Desafio DWBP	Desafio DWMM
Preservação	Preservar identificador do conjunto de dados quando o mesmo for descontinuado (removido ou arquivado) e prover informações sobre o arquivamento do recurso, além de possibilitar o acesso dos dados ao usuário.

Fonte: A autora (2019)

As Saídas esperados para a Dimensão Remoção, as boas práticas e os benefícios gerados por elas são descritos, nos três níveis de maturidade propostos pelo DWMM, no Quadro 23.

Levando em consideração os níveis de maturidade e os aspectos já mencionados a respeito da dimensão Remoção, uma visão do DWMM para a fase de Remoção pode ser vista no Quadro 24.

Quadro 23 – Dimensão Remoção do DWMM

REMOÇÃO			
Nível	Saídas	Boas Práticas Aplicadas	Benefícios
Ad-hoc	Não há preservação do identificador do conjunto de dados; Não é informado o motivo da indisponibilidade do conjunto de dados;	Nenhuma	Nenhum
Parcial	Não é realizada uma avaliação de cobertura do conjunto de dados; O <i>Documento de Solicitação de Remoção</i> é preenchido pelo <i>Provedor de Dados</i> com a informação do motivo de da remoção de acesso ao conjunto de dados;	BP22 - Fornecer uma explicação para os dados que não estão disponíveis	Reuso Confiabilidade
Definido	É realizada uma avaliação de cobertura antes da remoção de acesso ao conjunto de dados; É realizada a preservação dos identificadores do conjunto de dados; O consumidor é informado do motivo da remoção de acesso ao conjunto de dados;	BP22 - Fornecer uma explicação para os dados que não estão disponíveis BP27 - Preservar identificadores BP28 - Avaliar a cobertura do conjunto de dados	Reuso Confiabilidade

Fonte: A autora (2019)

Quadro 24 – Elementos da Dimensão Remoção

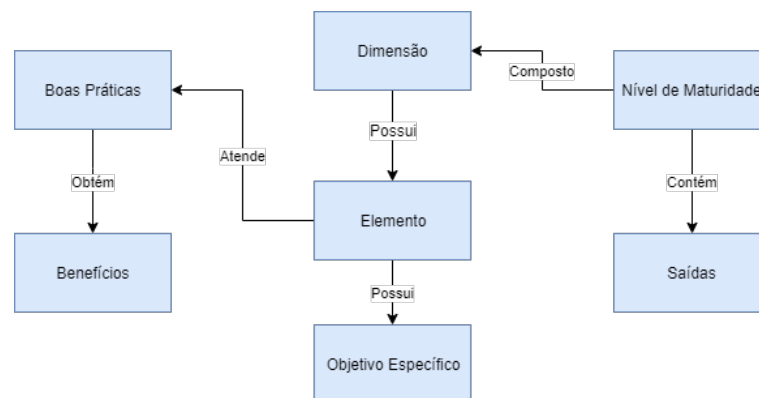
Remoção			
Elemento	Ad-hoc	Parcial	Definido
Preservação	O <i>Provedor de Dados</i> opta pela remoção de acesso ao conjunto de dados, mas não informa o motivo da remoção por meio do <i>Documento de Solicitação de Remoção</i> e o <i>Administrador</i> não realiza a preservação do identificador (URI).	O <i>Provedor de Dados</i> preenche o <i>Documento de Solicitação de Remoção</i> , informa o motivo e cria uma página de resposta informando ao consumidor que o conjunto não está mais disponível e qual o motivo da remoção.	É realizada uma avaliação de cobertura para que ao longo do tempo não seja perdido dados por falta de disponibilidade na Web.

Fonte: A autora (2019)

4.6 CONSIDERAÇÕES FINAIS

O DWMM possui um conjunto de 1 a 6 elementos para cada uma das 6 dimensões propostas. O modelo também utiliza, em todas as suas Dimensões, as boas práticas de publicação de dados na Web como um meio para que os atores possam enfrentar os desafios dos dados na Web que surgem durante cada fase do Ciclo de Vida de Dados na Web DWLM. Um panorama geral do modelo pode ser visto na Figura 19.

Figura 19 – Estrutura do DWMM.



Fonte: A autora (2019)

A partir da Figura é possível observar que:

- **Nível de Maturidade:** Contém Saídas, é composto de um ou mais elementos.
- **Dimensão:** Possui um ou mais elementos e Saídas ou Resultados Esperados para cada Nível de Maturidade.
- **Elemento:** Representa os desafios dos dados na Web que são enfrentados em cada Dimensão. Cada Elemento pode ser atendido/enfrentado por uma ou mais Boas Práticas e possui um Objetivo Específico em cada Dimensão do DWMM.
- **Objetivo Específico:** Descreve o propósito de cada Elemento ou Desafio em cada Dimensão do DWMM. Ou seja, embora os Elementos possam se repetir ao longo das Dimensões, eles possuem um propósito diferente em cada uma delas.
- **Boas Práticas:** Recomendações do grupo de trabalho do W3C para enfrentar os desafios dos dados na Web. Cada Boa Prática, ao ser aplicada, obtém um ou mais Benefícios dos dados na Web, que se tornam mais fortes na medida em que a adoção das boas práticas aumentam.
- **Benefícios:** Representa uma melhoria na maneira como os conjuntos de dados estão disponíveis na Web.

5 AVALIAÇÃO DO DWMM

Neste capítulo é discutida a avaliação do Modelo de Maturidade dos Dados na Web proposto neste trabalho. Para essa avaliação, foi utilizado o método qualitativo de Grupo Focal (*Focus Group*). Dessa forma, a Seção 5.1 apresenta o que constitui um grupo focal, enquanto a Seção 5.2 descreve sua organização, a análise dos dados (Seção 5.2.3). Por fim, na Seção 5.3 são apresentadas as considerações finais do Capítulo.

5.1 GRUPO FOCAL

Segundo Wilkinson e Silverman (2004), a pesquisa por meio de grupos focais, tradicionalmente, consiste em uma maneira de coletar dados qualitativos a partir do engajamento de um pequeno número de pessoas envolvidas em uma discussão informal, focadas em torno de um tópico particular ou um conjunto de questões. Nos últimos anos, o método de pesquisa por meio de grupos focais tem se tornado popular em diversas áreas de pesquisa, como Engenharia de Software, Medicina, Ciências Sociais, Biologia e ciências da Informação (ZAGANELLI et al., 2015).

Para Backes et al. (2011), o grupo focal como técnica de coleta e de análise de dados se constitui em uma importante estratégia para inserir os participantes da pesquisa no contexto das discussões de análise e síntese que contribuam para o repensar de atitudes, concepções, práticas e políticas sociais devido a sua capacidade interativa e problematizadora. Além disso, grupos focais podem ser efetivamente aplicados para promover o refinamento de um artefato proposto e a avaliação de sua utilidade (TREMBLAY; HEVNER; BERNDT, 2010).

Para Chiara (2005), o grupo focal não busca obter consenso, o moderador é que deve criar condições para que diferentes percepções e pontos de vista sejam colocados durante as sessões.

5.2 ORGANIZAÇÃO DO GRUPO FOCAL

Para a organização do Grupo Focal foram seguidas três fases estipuladas por Chiara (2005): planejamento, condução do grupo focal e análise dos dados. Essas três etapas serão detalhadas nas próximas seções.

5.2.1 Planejamento

Na etapa de planejamento são formuladas questões-chave de orientação à condução do grupo. A recomendação de Krueger (1994 apud CHIARA, 2005) é formular questões como: *Por que o estudo? Que informações deverão ser obtidas? Para quem elas serão úteis?*

Como localizar os participantes? Quais as categorias de participantes? Onde realizar as discussões? Quais as questões que serão feitas? Quem conduzirá as reuniões?

Para (CHIARA, 2005) o início do planejamento deve começar pela definição clara dos objetivos da realização do grupo focal, que devem ser coerentes com os objetivos da pesquisa, os critérios para selecionar os participantes, decidir o local de realização das discussões, a duração da sessão, as questões a serem feitas e elaborar os documentos a serem entregues aos participantes com os objetivos do estudo.

Deste modo, o objetivo principal do estudo foi realizar uma avaliação de alto nível a respeito da viabilidade, completude e adequação do Modelo de Maturidade para Dados na Web. Para essa avaliação inicialmente foram selecionados dez possíveis participantes que poderiam atender aos critérios estipulados. Esses critérios foram:

- Conhecimento e expertise em dados na Web (publicação e consumo de dados);
- Conhecimento e experiência em modelos de maturidade;
- Disposição em compartilhar suas experiências e opiniões;

Foi enviado um convite para os dez possíveis profissionais que, em nossa opinião, atenderiam aos critérios definidos. Após a realização dos dez convites, conseguimos conciliar a agenda de seis profissionais que se dispuseram a participar da pesquisa.

Na tabela 29 são apresentadas as informações gerais dos participantes deste estudo. Todos os participantes são residentes do estado de Pernambuco, Brasil. Quanto aos seus cargos, dois são estudantes (mestrado), onde um deles também desenvolve a função de Engenheiro de Testes. Há também um Professor e um Diretor de Governança de TI e Gestão de Processo, um Engenheiro de Software e um Técnico em Tecnologia da Informação. Dentre os seis, apenas um deles desenvolve atividades de dados na Web apenas no meio profissional; um desenvolve atividades apenas no meio acadêmico e os outros quatro em ambos os meios. Em relação à formação, dois possuem doutorado completo, um possui especialização, um possui mestrado completo e dois possuem graduação completa. Com exceção de um participante que tem apenas 1 ano de experiência em Dados na Web, todos os demais possuem ao menos 3 anos de experiência com Dados na Web. Já com relação à experiência com Modelos de Maturidade, apenas dois participantes afirmaram ter, no mínimo, 2 anos de experiência.

Por não residir na região metropolitana de Recife, um dos participantes se disponibilizou a participar remotamente da sessão. Para os demais, foi escolhido um local de comum acordo, diante disso o lugar escolhido foi o Centro de Informática da Universidade Federal de Pernambuco CIn/UFPE. A sessão foi realizada em Julho de 2019, com início às 9h e término às 11h35, ou seja, foram 2h35 de duração. Para a realização da sessão, foi escolhido um moderador, que ficou responsável por conduzir toda a sessão e assegurar que o foco de discussão estivesse dentro do tema abordado. Também foi definido um observador,

Tabela 29 – Informações dos Participantes do Grupo Focal

Participante	Função/ Cargo	Formação	Experiência com Dados na Web	Experiência com Mo- delos de Maturidade	Meio que desenvolve as atividades
PA1	Pesquisador / Estudante de Mestrado	Graduação	3 anos	Não	Acadêmico
PA2	Engenheiro de Testes / Estu- dante de Mes- trado	Graduação	6 anos	Não	Acadêmico e Profissional
PA3	Engenheiro de Software	Mestrado	4 anos	Não	Acadêmico e Profissional
PA4	Professor Uni- versitário	Doutorado	8 anos	3 anos	Acadêmico e Profissional
PA5	Diretor de Go- vernança de TI e Gestão de Processo	Doutorado	1 ano	2 anos	Profissional
PA6	Técnico em Tecnologia da Informação	Especialização	3 anos	Não	Acadêmico e Profissional

Fonte: A autora (2019)

que ficou responsável por realizar todas as anotações pertinentes durante a sessão e pela gravação de áudio da sessão.

5.2.2 Condução do Grupo Focal

A sessão do grupo focal começou com uma breve apresentação do moderador, observador e participantes. Cada um, antes do início das discussões, teve a oportunidade de se apresentar para os demais integrantes. Também foi entregue a cada participante o termo de consentimento para que o participante ficasse inteirado do que seria discutido e que teriam o seu anonimato garantido. Também foi informado que a sessão seria gravada e perguntado se todos estavam de acordo. Juntamente com o termo, também foi entregue um questionário a cada participante. Esse questionário continha as informações que deveriam ser avaliadas por cada um a respeito dos principais elementos do DWMM.

Na sequência, o moderador abriu a sessão com uma breve explicação da pesquisa, descrevendo o assunto que seria discutido e os objetivos do grupo focal. Como todos os participantes tinham experiência na área de Dados na Web, foi introduzido, de uma forma geral, o conceito de Modelos de Maturidade para que todos os participantes tivessem uma

noção básica do tema abordado. Além disso, também foram apresentadas algumas regras de funcionamento:

- Nós queremos ouvir a opinião de todos;
- Todos os comentários são válidos;
- Não existem respostas corretas ou incorretas;
- Não fugir de tópico.

Logo após, foi explicado que o grupo focal seria dividido em três fases. Na primeira, os participantes iriam preencher as suas informações básicas (*i.e* função/cargo, formação, experiência, meio que desenvolve as atividades). A segunda seria um questionário com vinte e duas perguntas a respeito dos componentes (em conjunto e separado) do DWMM. E por último, na terceira fase seriam duas perguntas gerais, a fim de coletar dados de avaliação de qualidade, bem como comentários gerais. O formulário aplicado pode ser encontrado no Apêndice A.

5.2.3 Análise dos Dados

Esta seção apresenta a análise dos dados coletados no grupo focal, discutindo detalhadamente cada pergunta realizada e apontando algumas questões que devem ser levadas em consideração.

5.2.3.1 Avaliação dos Níveis de Maturidade

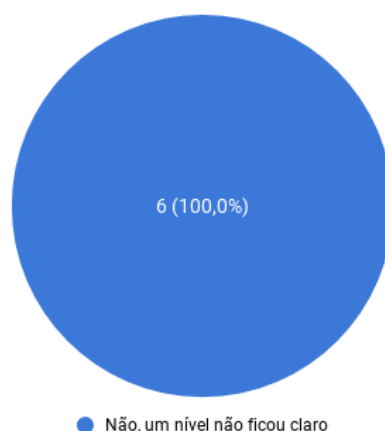
O primeiro conjunto de perguntas foi relacionado aos níveis de maturidade propostos no DWMM. Primeiramente, foi perguntado se os níveis de maturidade eram fáceis de serem interpretados. Com esse questionamento, almejávamos detectar a existência de lacunas, erros e/ou possibilidades de melhoria. A resposta dessa pergunta foi unânime (ver Figura 20), todos os participantes selecionaram “Não, um nível não ficou claro”.

PA1 informou “*Explicar melhor o nível 2. Não está claro do que se trata ou suas diferenças do nível 3*”. PA2 afirmou “*Os níveis poderiam ser melhor detalhados*”.

PA3 afirmou “*Há uma diferença muito grande entre o nível 2 e o nível 3*”. Já o PA4 expôs uma opinião mais quanto aos nomes dos níveis do que quanto à descrição geral de cada nível em si, “*Acho que o título dos níveis está confuso. Por exemplo, o nível otimizado não representa atividades otimizadas. Talvez, esteja mais próximo de definido. E o nível moderado talvez esteja mais próximo de parcial. Eu acho que deveria ter mais níveis intermediários. Até para permitir a monitoração e otimização dos processos...*”.

PA5 comentou “*A descrição do nível 2 especifica apenas que o ciclo de vida e as boas práticas são conhecidas. Acrescentaria que algumas ou nenhuma boa prática são aplicadas. O termo “Moderado” requer um certo nível de aplicabilidade e se uma organização*

Figura 20 – Avaliação dos Níveis do DWMM.

Avaliação dos Níveis de Maturidade do DWMM

Fonte: A autora (2019)

apenas conhece, sem aplicar nada, não consideraria moderado. (viabilidade, completude, adequação)". PA6 também comentou que "Entre os níveis moderado e inexistente ficou incerta a ideia do que seriam "algumas práticas" e "todas as práticas"".

Com as melhorias propostas identificamos que não ficaram claras as descrições dos níveis definidos para o DWMM. A partir disso, detalhamos mais as descrições de cada nível, a fim de cobrir essa lacuna.

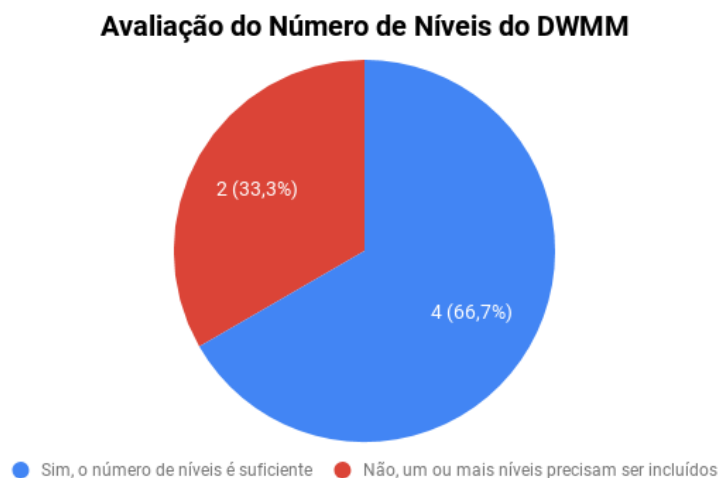
Também percebemos que o PA4 fugiu um pouco da resposta para a pergunta sobre as descrições dos níveis, e que o mesmo teve dúvidas com relação ao objetivo do modelo, em partes porque o mesmo estava com modelos de capacidade em mente e visando aspectos políticos e organizacionais, o que nos fez frisar ao longo do texto que este modelo tem representação por estágios, ou seja, o foco dele não é avaliar a maturidade de um único processo (ou elemento), mas de avaliar a maturidade do conjunto de elementos para cada Dimensão, ou seja, para avaliar a Dimensão Planejamento, precisamos avaliar o conjunto de elementos definidos para a Dimensão. E a avaliação da Dimensão Criação, só é possível após o resultado da avaliação da Dimensão Planejamento. Ademais, embora não tenha sido o escopo da questão, foi atendido o comentário com as sugestões para os nomes dos níveis que o PA4 comentou.

A segunda pergunta sobre os níveis propostos pelo DWMM foi se o número de níveis era suficiente. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria. A Figura 21 resume as respostas.

Os participantes PA1 e PA2 selecionaram "Sim, o número de níveis é suficiente" e não fizeram nenhum comentário adicional.

O PA3 selecionou "Não, um ou mais níveis precisam ser incluídos" e comentou "*Pode ser adicionado um novo nível entre o nível 2 e 3, com o foco principal na aplicação parcial*

Figura 21 – Avaliação do Número de Níveis do DWMM.



Fonte: A autora (2019)

das boas práticas. Em alternativa, o nível 2 pode ser complementado para abranger este foco”.

O PA4 selecionou “Não, um ou mais níveis precisam ser alterados” e comentou “A quantidade de níveis não permite uma implantação de uma forma incremental... Está faltando níveis intermediários. Outra coisa, está faltando quais são os resultados esperados para cada um dos níveis...”.

O PA5 selecionou “Sim, o número de níveis é suficiente” e comentou “Eu gosto da ideia de 3 níveis mais simples e ágil. Porém, necessário fazer ajustes como “conhecido”, “otimizado”, “aplicado”. Senti falta de mensurar o resultado das práticas adotadas. Algumas práticas podem dar a ideia de 2 ou 33 práticas. A precisão do termo é muito baixa”.

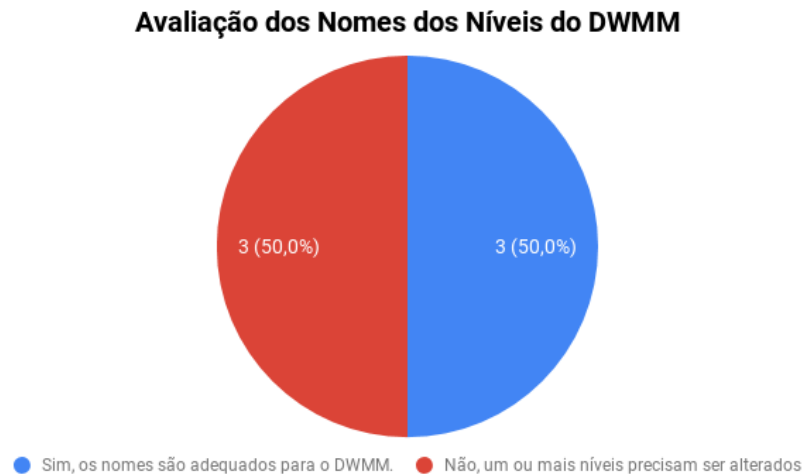
O PA6 selecionou “Sim, o número de níveis é suficiente” e informou “Não é necessário a inclusão de mais níveis. Talvez seja preciso ajustar os nomes dos níveis de forma a termos uma noção mais “mensurável” do que se trata o nível”.

Apenas os participantes PA3 e PA4 afirmaram sobre a necessidade da inclusão de mais níveis, entretanto o PA3 sugeriu, também, uma alteração no nível 2 para que não seja necessário a inclusão de outro nível. Quanto ao comentário do PA4, a respeito de pontuar os resultados esperados para cada um dos níveis, atendemos isso por meio da realização de um Quadro explicativo com a visão geral de cada Dimensão, onde pontuamos as Saídas para cada nível, as boas práticas que são aplicadas e os benefícios gerados por meio da adoção das boas práticas. Desta forma, o número de níveis propostos pelo DWMM foi mantido em 3.

A última pergunta com relação aos níveis de maturidade foi sobre os nomes propostos. Durante a proposta do DWMM, tivemos incerteza quanto aos nomes que estávamos propondo e decidimos realizar uma pergunta a fim de identificar se os nomes eram adequados

e receber sugestões de melhorias. A Figura 22 resume as respostas.

Figura 22 – Avaliação do Número de Níveis do DWMM.



Fonte: A autora (2019)

Os participantes *PA1*, *PA2* e *PA3* selecionaram “Sim, os nomes são adequados para o DWMM” e não fizeram nenhum comentário adicional. Os demais selecionaram “Não, um ou mais níveis precisam ser alterados”.

O *PA4* comentou “*Acho que os níveis não fazem sentido quando analisamos que o grão menor não é um conjunto de dados. É o projeto de publicação de dados, que trata de mais de um conjunto de dados. Assim, um nível gerenciado e otimizado faz muito sentido*”.

O *PA5* comentou “*Acho que o nível 3 não seria “otimizado”. Otimizado remete a algo que foi mensurado/medido. No caso, não bastaria adotar todas as boas práticas, mas medi-las e avaliar o resultado*”.

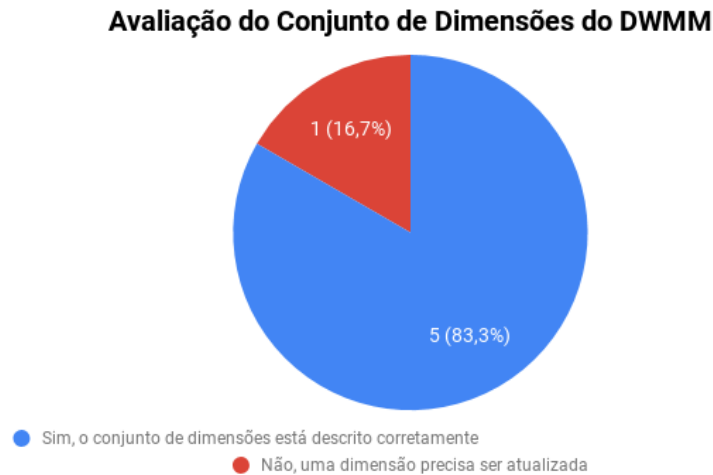
O *PA6* reiterou o que já havia comentado na resposta anterior, “*...Talvez seja preciso ajustar os nomes dos níveis de forma a termos uma noção mais “mensurável” do que se trata o nível*”.

Mais uma vez, houve uma falta de entendimento do *PA4* quanto ao objetivo do modelo, que será discutido nas questões mais à frente. A partir dos comentários recebidos, alteramos os nomes de *Inexistente* para *Ad-hoc*, *Moderado* para *Parcial* e *Otimizado* para *Definido*.

5.2.3.2 Avaliação do Conjunto de Dimensões do DWMM

A quarta questão foi relacionada às dimensões do DWMM de uma maneira geral e teve como objetivo identificar se elas foram descritas corretamente. Com isso, tínhamos o objetivo de identificar lacunas, erros e/ou possibilidades de melhoria. A Figura 23 resume as respostas.

Figura 23 – Avaliação do Conjunto de Dimensões do DWMM.



Fonte: A autora (2019)

PA2 selecionou “Não, uma dimensão precisa ser atualizada” e comentou “*Rever o nome Remoção, talvez o nome não esteja adequado*”.

PA4 selecionou “Sim, o conjunto de dimensões está descrito corretamente” mas comentou “*As dimensões estão bem descritas e fazem sentido. Mas, não sei se é a melhor organização. Será que não deveria ser de outra forma? Minha questão é que não sei se dividir dimensões por fases é a melhor opção. Será que por problema não era melhor? Tipo gerenciamento de metadados, gerenciamento de infraestrutura, gerenciamento de refinamento?...*”.

PA5 selecionou “Sim, o conjunto de dimensões está descrito corretamente” e comentou “*Dúvida apenas no Remoção. Remover acesso e visualização? Não está mais publicado? Nem para visualização nem para acesso?*”

Os demais participantes selecionaram “Sim, o conjunto de dimensões está descrito corretamente” e não pontuaram nada em relação a pergunta.

Quanto aos comentários do PA2, não concordamos, visto que estamos levando em consideração as etapas do Modelo de Ciclo de Vida dos Dados na Web - DWLM. Inicialmente, o DWLM definiu o nome desta fase como *Arquivamento*, no entanto, ao passar por uma avaliação de Grupo Focal, o nome da fase foi alterado para *Remoção*. Desta forma, decidimos continuar com o nome *Remoção* para a Dimensão do DWMM.

Mais uma vez, o PA4 expôs comentários quanto a questões organizacionais e de processos. No entanto, nossa escolha foi de utilizar as boas práticas que abordam apenas aspectos técnicos e não políticos e organizacionais. O comentário dele não se aplica ao cenário do DWMM, pois o objetivo do modelo é avaliar tanto organizações quanto indivíduos, e no caso mencionado, um indivíduo não teria o mesmo cenário de infraestrutura de uma organização.

5.2.3.3 Avaliação da Dimensão Planejamento

Após a avaliação do Conjunto de Dimensões, foi iniciada a avaliação de cada uma das dimensões. Os participantes foram questionados quanto à descrição da Dimensão (ver Figura 24), se os objetivos e os três níveis de maturidade para esta dimensão haviam sido definidos de forma clara. Quanto à descrição, almejávamos identificar se os elementos eram viáveis, completos e adequados. Quanto aos objetivos, também almejávamos identificar se os objetivos de cada nível estavam claramente definidos, se eram completos e se podiam ser alcançados. E por fim, quanto aos níveis, almejávamos identificar se os níveis estavam descritos de forma clara e concisa para cada Elemento da Dimensão e se eram completos e podiam ser alcançados, bem como detectar a existência de lacunas, erros e/ou possibilidades de melhoria.

Figura 24 – Avaliação da Dimensão Planejamento.

Avaliação Dimensão Planejamento do DWMM



Fonte: A autora (2019)

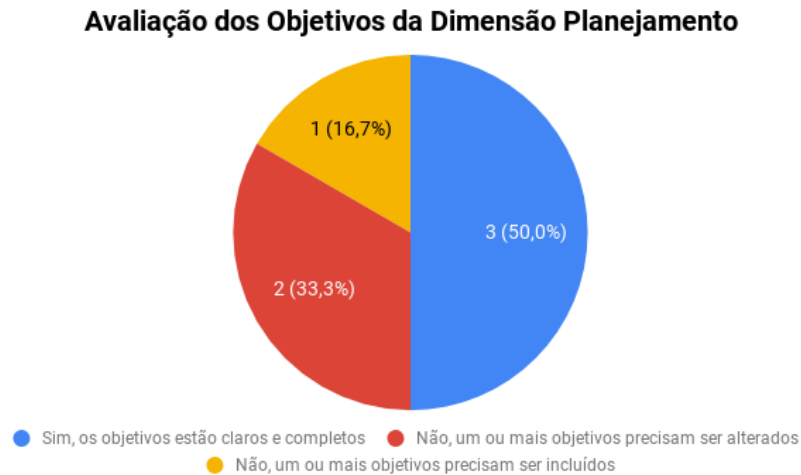
Para o primeiro questionamento, os participantes *PA1*, *PA2* e *PA6* selecionaram “Sim, a dimensão está descrita corretamente” e não fizeram nenhum comentário adicional.

PA3 selecionou “Não, um ou mais elementos precisam ser atualizados” e comentou “Fiquei em dúvida se o elemento de qualidade estaria aqui nesta dimensão, ou se qualidade é um elemento de fato, já que a qualidade dos dados pode ser também a consequência da aplicação do modelo de maturidade”.

PA4 selecionou “Não, um ou mais elementos precisam ser incluídos” e comentou “Senti falta de elementos que planejam a infraestrutura... De uma forma geral, acredito que está faltando como elemento estrutural do modelo de maturidade uma listagem de outcomes para cada nível. Isso permite uma melhor medição se o nível de maturidade foi alcançado”.

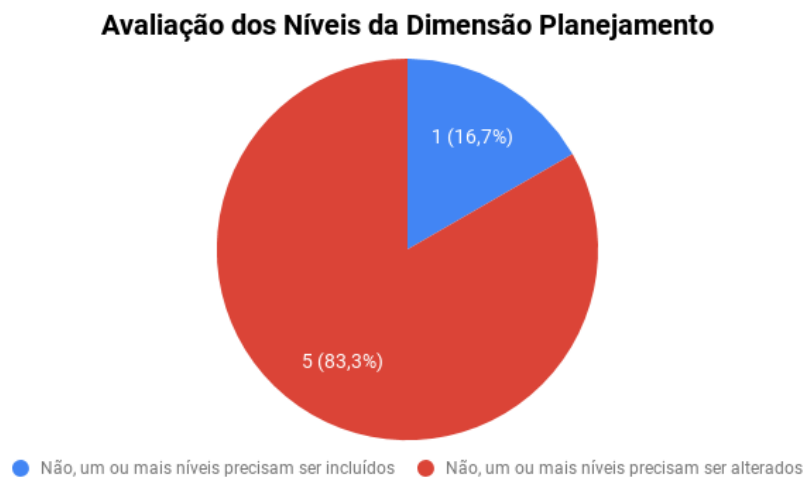
PA5 selecionou “Não, um ou mais elementos precisam ser incluídos” e comentou “Senti falta na descrição da dimensão contemplar definições referentes as outras fases do ciclo de

Figura 25 – Avaliação dos Objetivos da Dimensão Planejamento.



Fonte: A autora (2019)

Figura 26 – Avaliação dos Níveis da Dimensão Planejamento.



Fonte: A autora (2019)

vida, vejo muito criação e publicação. Sinto falta de definição referentes a outras etapas. O planejamento remete a definições em todas as fases”.

Para essas sugestões, não concordamos com a opinião dos participantes. Com relação ao comentário do participante *PA3*, o elemento qualidade é um dos desafios que são enfrentados na etapa de planejamento. Diz respeito aos critérios de qualidade que são preenchidos ou não no *Documento de Descrição do Conjunto de Dados*. Concordamos que, de fato, a qualidade dos dados pode ser também a consequência da aplicação do modelo de maturidade, à medida que você o segue e aplica as boas práticas informadas. No entanto, não estamos avaliando a qualidade dos dados nessa fase, mas se o *Provedor de Dados* definiu os critérios de qualidade a serem utilizados para a avaliação do conjunto

na fase de Criação.

Com relação ao comentário do participante *PA4*, mais uma vez foi mencionado aspectos organizacionais, que não são considerados nesse trabalho. No entanto, a sugestão da listagem de saídas para cada nível foi atendida. Já o participante *PA5* comentou sobre a fase de planejamento do DWLM, uma vez que o DWMM utiliza as fases do DWLM, não é possível atender a sugestão do participante *PA5*, pois deveria ser alterada a forma como a fase de Planejamento é definida no DWLM, o que não vem ao caso.

Para o segundo questionamento (ver Figura 25), os participantes *PA1*, *PA2* e *PA3* selecionaram “Sim, os objetivos estão claros e completos” e não fizeram nenhum comentário adicional.

PA4 selecionou “Não, um ou mais objetivos precisam ser alterados” e comentou “*Parece que outras questões relacionadas a planejamento estão faltando. Como disse na resposta anterior, tem uma série de outros problemas que está faltando... Analise todos os objetivos para ver se eles continuam fazer sentido com mais de um conjunto de dados*”. Quanto ao comentário feito pelo participante *PA4*, para avaliar a maturidade de uma organização ou indivíduo no processo de publicação de dados, foi utilizado o processo que foi adotado para publicação de um ou mais conjuntos de dados, no caso o DWLM. Nesse caso, a avaliação do DWMM é feita por conjunto de dados.

PA5 selecionou “Não, um ou mais objetivos precisam ser incluídos” e reiterou o comentário anterior com relação às outras fases, como consumo, refinamento e remoção na Dimensão Planejamento. O motivo pelo qual isso não ocorre é que não há como avaliar a forma em que o conjunto de dados será consumido, desta forma, não há como ter um planejamento do que acontece durante e após a fase de consumo.

PA6 selecionou “Não, um ou mais objetivos precisam ser alterados” e comentou “*Seria interessante esclarecer uma justificativa do porquê dos elementos que foram escolhidos*”. No entanto, essa justificativa já havia sido dada no Capítulo 4. Os elementos foram identificados a partir das atividades definidas no DWLM.

O último questionamento desta dimensão foi sobre os níveis de maturidade para cada elemento (ver Figura 26), se haviam sido claramente definidos a fim de identificar lacunas, erros e/ou possibilidades de melhoria.

PA1 selecionou “Não, um ou mais níveis precisam ser incluídos”, os demais participantes selecionaram “Não, um ou mais níveis precisam ser alterados”.

PA1 comentou “*Descrever melhor as diferenças do n2 e n3 em metadados. Em qualidade, não há diferenças entre n1 e n2*”. Não há diferenças entre o nível 1 e 2 do Elemento Qualidade, pois definimos que uma organização ou indivíduo só é maduro o suficiente para avaliar o conjunto de dados quanto à sua qualidade no nível 3, visto que é uma tarefa complexa a se fazer. Não faz sentido definir os critérios de qualidade no nível 2 da Dimensão Planejamento e apenas avaliar a qualidade no nível 3 da Dimensão Criação. Desta forma, não aceitamos a sugestão do participante *PA1*.

PA2 comentou “Melhorar a descrição de cada nível x elemento incluindo as limitações, quando houver”. Já o *PA3* comentou “Particularmente, não gostei muito de ver alguns quadrados no nível 3 em branco. Poderia ser colocado alguma informação que indique a maturidade alcançada”. O participante *PA6* também reiterou “Evitar deixar a descrição de um nível em branco”. Essas sugestões de melhorias foram atendidas. Para os níveis que estavam em branco, repetimos a descrição do nível anterior, uma vez que estamos avaliando o conjunto de elementos nos três níveis, realmente não faria sentido avaliar o nível 3 com a descrição em branco. Desta forma, melhoramos as descrições e para cada Elemento com nível em branco, foi repetida a descrição do nível anterior.

PA4 comentou “Alguns níveis estão com uma descrição meio que booleana. Falta níveis intermediários. Ou mesmo uma melhor diferenciação entre os níveis”. De fato, algumas descrições ficaram booleanas, pois estamos vendo se a informação existe ou não, por se tratar da fase de Planejamento.

PA5 ficou com dúvida em relação aos critérios especificados e comentou “...São aplicáveis para todo tipo de conjunto de dados? Ex. metadados, nível 3. Todos os elementos ali são aplicáveis em todos os tipos de conjuntos de dados?”. A resposta é sim, todos os critérios especificados para os metadados descritivos e estruturais contidos no nível 3 são aplicáveis para todos os conjuntos de dados.

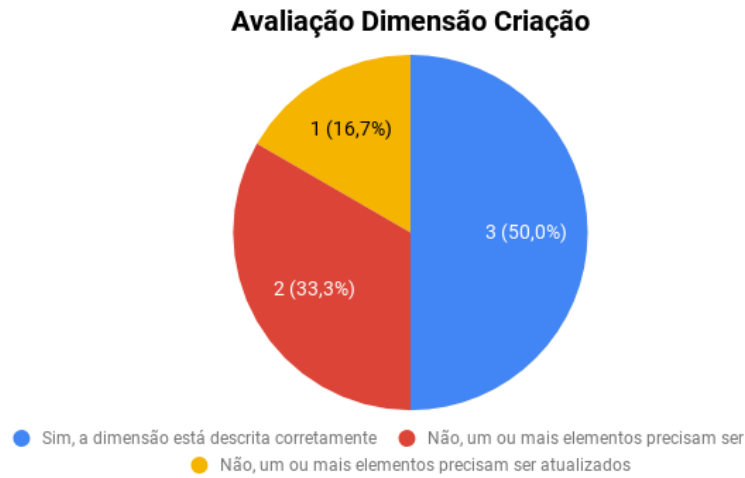
5.2.3.4 Avaliação da Dimensão Criação

De forma semelhante à Dimensão Planejamento, também avaliamos se a descrição da dimensão (ver Figura 27), os objetivos e os três níveis de maturidade para a Dimensão Criação haviam sido definidos de forma clara. Quanto à descrição, almejávamos identificar se os elementos eram viáveis, completos e adequados. Quanto aos objetivos, também almejávamos identificar se os objetivos de cada nível estavam claramente definidos, se eram completos e se podiam ser alcançados. E por fim, quanto aos níveis, almejávamos identificar se os níveis estavam descritos de forma clara e concisa para cada elemento da dimensão e se eram completos e podiam ser alcançados, bem como detectar a existência de lacunas, erros e/ou possibilidades de melhoria.

Os participantes *PA1*, *PA2* e *PA3* selecionaram “Sim, a dimensão está descrita corretamente” e não fizeram nenhum comentário adicional.

PA4 selecionou “Não, um ou mais elementos precisam ser incluídos” e comentou “Não deveria ter um desafio voltado à produção ou extração dos dados? De onde os dados virão? Em relação a proveniência. Por exemplo, informações sobre como os dados foram criados...”. Quanto aos dados de proveniência das fontes de dados, estão contidos no Documento de Descrição do Conjunto de Dados já na fase de Planejamento. A partir do comentário do *PA4*, acrescentamos a informação de proveniência sobre o processo de criação do conjunto de dados como *Outcome* do nível 3 da Dimensão Criação.

Figura 27 – Avaliação da Dimensão Criação.



Fonte: A autora (2019)

Figura 28 – Avaliação dos Objetivos da Dimensão Planejamento.

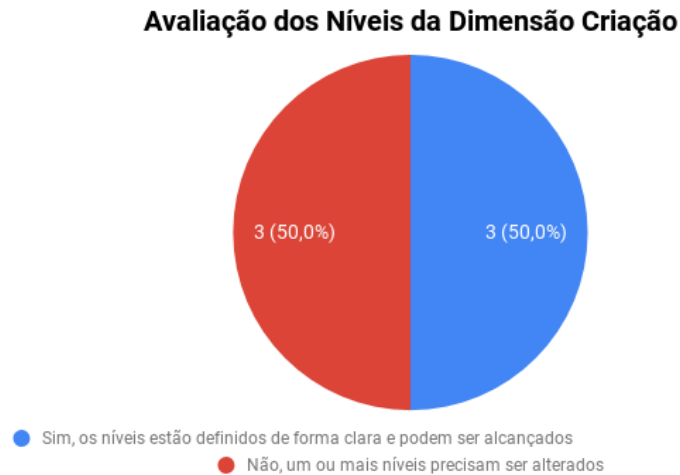


Fonte: A autora (2019)

PA5 selecionou “Não, um ou mais elementos precisam ser incluídos” e comentou “*Senti falta na descrição um texto sobre vocabulário*”. Da mesma forma, o PA6 comentou “*Rever os níveis do elemento vocabulário*”. Os comentários referentes ao Elemento Vocabulário serão realizados ao fim desta Seção.

Quanto aos objetivos da Dimensão com exceção do PA4, todos os participantes selecionaram “Sim, os objetivos estão claros e completos” (ver Figura 28). O PA2 pontuou “*Explicar melhor o elemento de vocabulário, pois não parece estar incluído nas atividades da fase*”. Já o PA4 comentou “*Em alguns desafios, não há progressão dos objetivos entre os níveis. Por exemplo, em proveniência deveria ser: nível 1 - não há informações de proveniência. Nível 2 - há informações a respeito da fonte e do produtor. Nível 3 - há*

Figura 29 – Avaliação dos Níveis da Dimensão Planejamento.



Fonte: A autora (2019)

informações a respeito da produção dos dados". Quanto ao comentário do *PA4* levamos em consideração e alteramos os níveis para abranger os pontos mencionados. A respeito do comentário do *PA2*, levamos em consideração e, juntamente com os comentários da questão anterior, decidimos excluir o elemento Vocabulário, visto que não há menção dele nas atividades do DWLM.

Por fim, foi perguntado a respeito dos níveis para a Dimensão Criação (ver Figura 29). Os participantes *PA1*, *PA2* e *PA6* selecionaram "Sim, os níveis estão definidos de forma clara e podem ser alcançados". Os participantes *PA3*, *PA4* e *PA5* selecionaram "Não, um ou mais níveis precisam ser alterados".

PA1 comentou "*Para todo o documento, é melhor encontrar um padrão nos níveis buscando não deixar níveis sem descrição. Talvez alocar descrições do n1 para n2 e n2 para n3*". Esta sugestão já foi realizada para todo o documento. O participante *PA3* comentou "*Rever o elemento vocabulário, pois só passa a existir no nível 3*". Como mencionado anteriormente, decidimos excluir o Elemento Vocabulário.

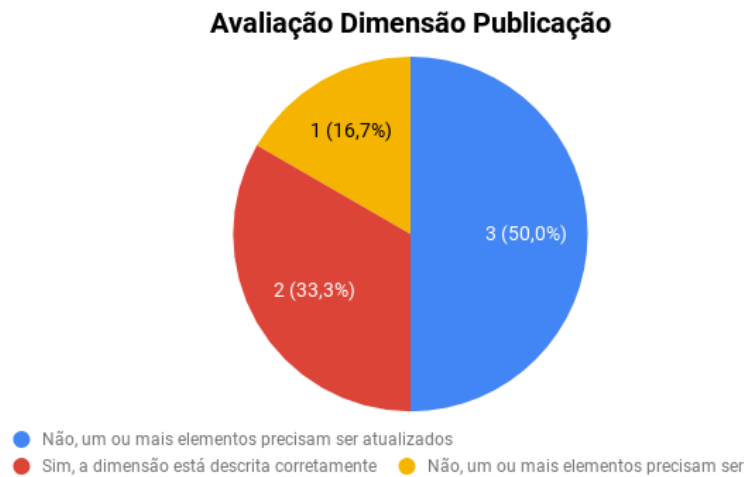
PA4 comentou "*Também acho estranho tanto ter níveis vazios quanto ter níveis em que não há evolução entre eles... É importante em um modelo de maturidade ter um mesmo padrão de evolução entre os níveis...*". A respeito dos níveis vazios, foi alterado para todas as dimensões e já foi comentado o motivo do fato de um mesmo elemento ter descrições iguais para níveis diferentes. Estamos avaliando o conjunto de Elementos da Dimensão em cada nível e não um elemento sozinho.

Do mesmo modo, o participante *PA5* comentou "*Acho estranho ter 2 níveis com mesmo nível de restrição. Ex. não faz uso de vocabulários*". A respeito do elemento, o mesmo já foi excluído e quanto ao mesmo nível de restrição, já foi explicado.

5.2.3.5 Avaliação da Dimensão Publicação

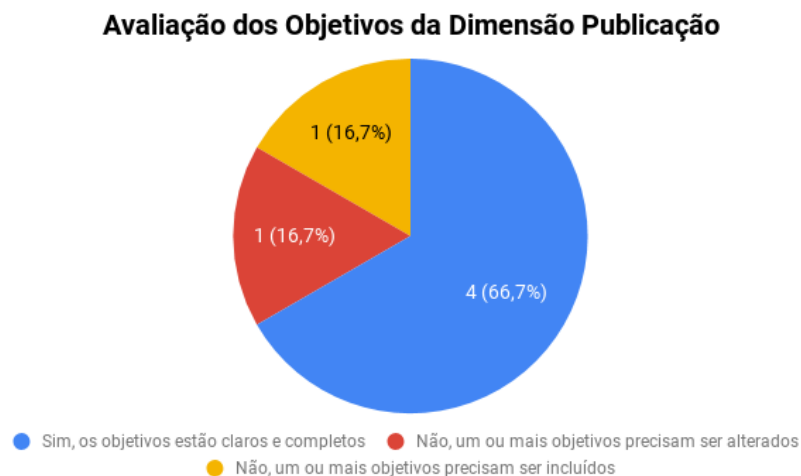
Assim como as Dimensões anteriores, também perguntamos quanto à descrição, objetivos e níveis da Dimensão Publicação. Pretendíamos identificar a existência de lacunas, erros e/ou possibilidades de melhorias.

Figura 30 – Avaliação da Dimensão Publicação.



Fonte: A autora (2019)

Figura 31 – Avaliação dos Objetivos da Dimensão Publicação.



Fonte: A autora (2019)

Quanto à descrição da Dimensão Publicação os participantes *PA2* e *PA6* selecionaram “Sim, a dimensão está descrita corretamente” e não fizeram nenhum comentário adicional. Os demais selecionaram “Não, um ou mais elementos precisam ser atualizados” (ver Figura 30).

Figura 32 – Avaliação dos Níveis da Dimensão Publicação.



Fonte: A autora (2019)

PA1 comentou “Mudar o nome “enriquecimento”. A definição da palavra talvez não sirva para o elemento”. Do mesmo modo, o PA3 comentou “Acredito que o nome do elemento Enriquecimento não se encaixa com a descrição do elemento”. No entanto, de acordo com (LÓSCIO; BURLE; CALEGARI, 2016), no desafio de Enriquecimento há a relação com a BP32 de prover informações complementares, o que está diretamente relacionado com a definição descrita para o Elemento no DWMM. Desta forma, não aceitamos os comentários dos participantes.

PA4 comentou “Os níveis da dimensão estão escritos como seguindo o modelo 5 estrelas. Isso é bem diferente das dimensões anteriores. Além disso, se licenças já foram definidas no nível anterior porque elas estão sendo tratadas aqui novamente? Uma vez que você alcançou isso não era para tratar em mais de um lugar. A incidência de múltiplos desafios cross-cutting torna difícil a implantação e avaliação”. Com relação à referência das 5 estrelas, não foi nossa intenção, por isso, foi ajustado o texto para que não contemplasse essa definição. A respeito da incidência de múltiplos desafios, cada desafio é tratado de uma maneira diferente em cada Dimensão.

PA5 comentou “A descrição contempla aspectos como de publicação, ficou faltando legível por humanos e máquinas”. No entanto, essa descrição está mencionada nos respectivos elementos.

Quanto aos objetivos da Dimensão os participantes PA1, PA2, PA3 e PA6 selecionaram “Sim, os objetivos estão claros e completos” e não fizeram nenhum comentário adicional (ver Figura 31).

O PA4 reiterou o comentário anterior “Pelo que eu entendi, o grande objetivo desta dimensão é alcançar o nível 5 estrelas de Berners-Lee. Mas, boa parte da forma de alcançar isso, a meu ver, é realizado pelas dimensões anteriores”. O texto foi ajustado para que esta

interpretação não ocorresse. E o participante *PA5* fez o mesmo comentário da pergunta anterior.

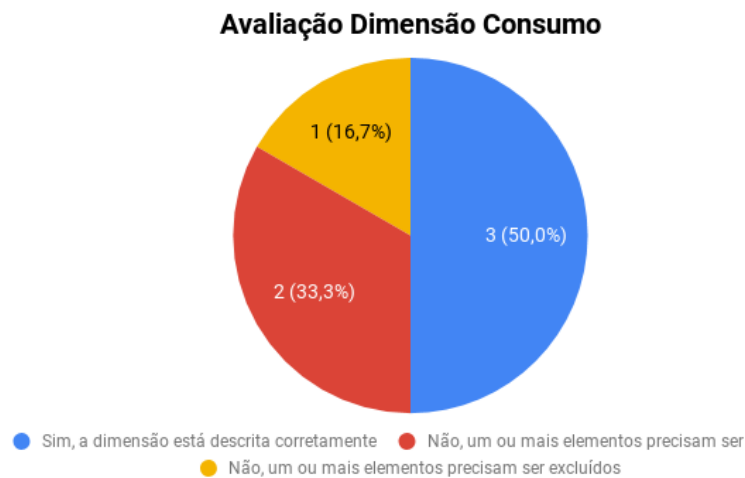
Por fim, quanto aos níveis, os participantes *PA2*, *PA3* e *PA6* selecionaram “Sim, os níveis estão definidos de forma clara e podem ser alcançados” e não fizeram mais nenhum comentário adicional (ver Figura 32).

PA1 comentou “*Definir ‘formato de difícil reutilização’ em ‘formato’*”. O texto foi alterado para o Elemento Formato da Dimensão Criação. O participante *PA4* comentou “*Novamente, não está claro o que quer ser alcançado em cada nível*”. Acreditamos que com a reformulação do Quadro 11 essa questão tenha sido resolvida. E o participante *PA5* repetiu o comentário da questão anterior.

5.2.3.6 Avaliação da Dimensão Consumo

Em sequência, foi realizado as mesmas três perguntas para a Dimensão Consumo, a fim de identificarmos lacunas, erros e/ou possibilidades de melhorias. Os participantes *PA1*, *PA3* e *PA5* selecionaram “Sim, a dimensão está descrita corretamente” e não pontuaram nada. A Figura 33 contém um resumo das respostas.

Figura 33 – Avaliação da Dimensão Consumo.

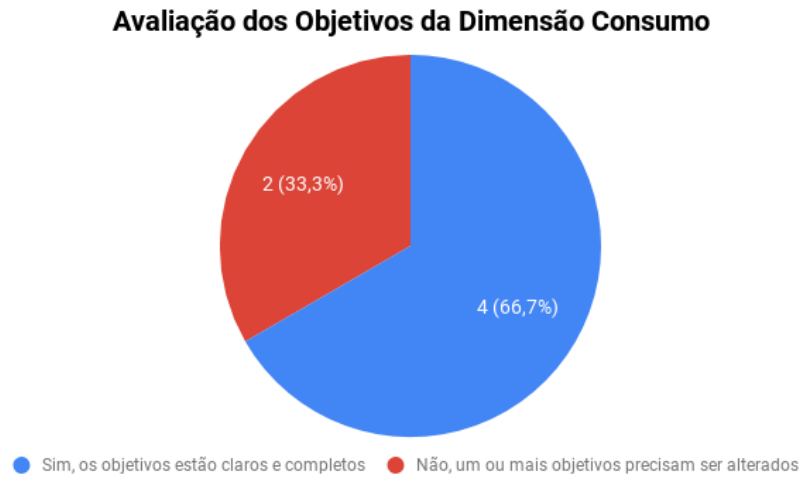


Fonte: A autora (2019)

Os participantes *PA2* e *PA6* selecionaram “Não, um ou mais elementos precisam ser incluídos”. O participante *PA2* comentou “*Adicionar elemento que meça o número de acessos e downloads aos conjuntos de dados*”. De igual modo o participante *PA6* comentou “*Rever alguns desafios para essa fase, como por exemplo acesso*”. Não há desafio ou Boa Prática que se relacione com essas sugestões, da mesma forma, não há atividade no DWLM que justifique a inclusão solicitada.

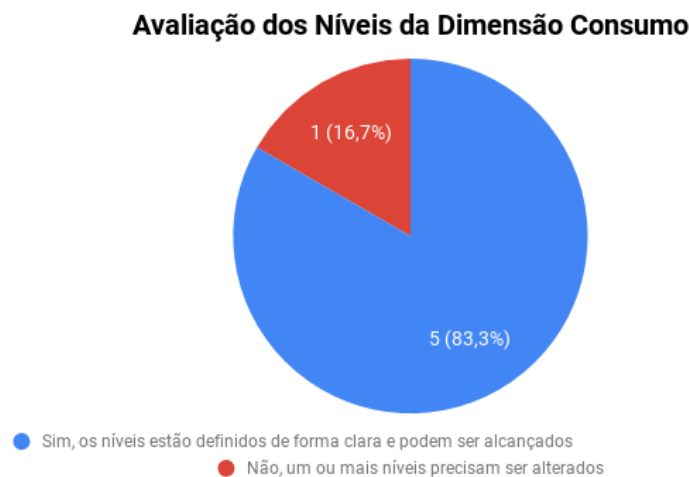
PA4 comentou “*Não faz sentido essa definição desta dimensão. Como você vai avaliar o consumo? Você está tratando apenas o feedback. Só.*”. Não estamos avaliando a fase

Figura 34 – Avaliação dos Objetivos da Dimensão Consumo.



Fonte: A autora (2019)

Figura 35 – Avaliação dos Níveis da Dimensão Consumo.



Fonte: A autora (2019)

Consumo, mas a Dimensão Consumo, e ela não mensura as atividades realizadas pelo consumidor.

A respeito dos objetivos da Dimensão Consumo os participantes *PA1*, *PA2*, *PA3* e *PA5* selecionaram “Sim, os objetivos estão claros e completos”. Apenas o *PA5* comentou “*Talvez incluir outros elementos referente ao consumo. Achei feedback pouco para a grandeza da fase*” (ver Figura 34).

PA4 selecionou “Não, um ou mais objetivos precisam ser alterados” e comentou “*O nível 3 não deveria fazer uso otimizado do feedback? Seria muito importante usar o conhecimento do consumidor de uma forma efetiva*”. E o participante *PA6* pontuou “*Explicar essa visão do consumo por meio do consumidor*”.

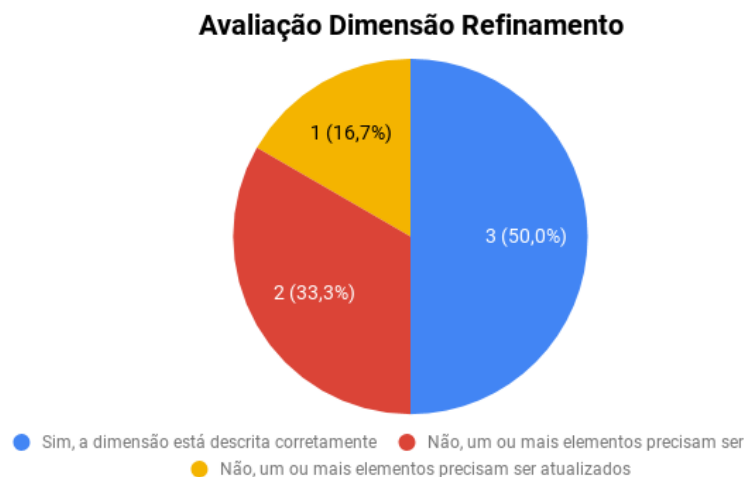
Quanto aos níveis (ver Figura 35), com exceção do *PA4*, todos os participantes selecionaram “Sim, os níveis estão definidos de forma clara e podem ser alcançados” e não fizeram nenhum comentário adicional. O *PA4* selecionou “Não, um ou mais níveis precisam ser alterados” e comentou “*Novamente, não é claro entender a evolução entre os níveis*”.

Além da clareza na evolução dos níveis, não concordamos com as demais sugestões dos participantes, pois não estamos avaliando a fase de Consumo do ponto de vista do consumidor, mas do publicador.

5.2.3.7 Avaliação da Dimensão Refinamento

Seguindo a mesma ideia, questionamos quanto à descrição, objetivos e níveis da Dimensão Refinamento a fim de identificarmos lacunas, erros e/ou possibilidades de melhorias. Para a descrição da Dimensão (ver Figura 36) os participantes *PA1*, *PA2* e *PA6* selecionaram “Sim, a dimensão está descrita corretamente” e não fizeram nenhum comentário adicional.

Figura 36 – Avaliação da Dimensão Refinamento.

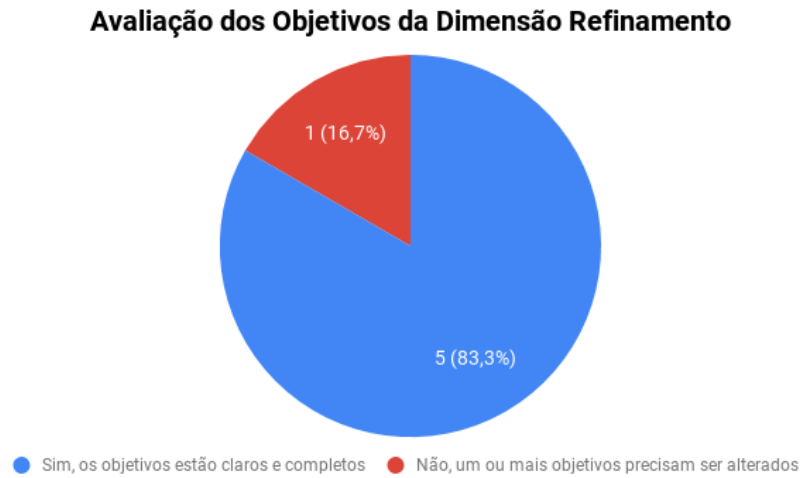


Fonte: A autora (2019)

PA3 selecionou “Não, um ou mais elementos precisam ser incluídos” e comentou “*Acredito que deva existir um elemento de Limpeza e um de Enriquecimento separadamente, pois são conceitos distintos*”. Como o nosso foco é avaliar como os desafios dos dados na Web são enfrentados ao longo do ciclo de vida dos dados na Web, não há como fazer a separação do Elemento Enriquecimento para a criação de um novo Elemento Limpeza.

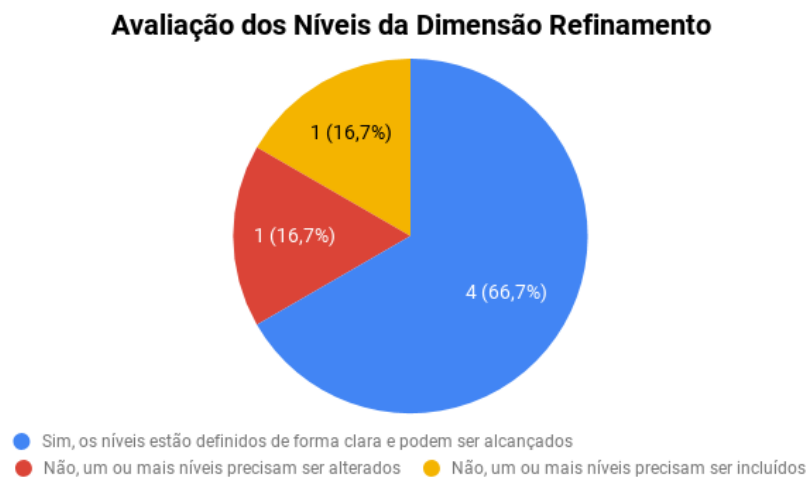
PA4 selecionou “Não, um ou mais elementos precisam ser atualizados” e comentou “*Não entendi por que versionamento entrou nesta dimensão. E, deveria ter um elemento limpeza de conjunto de dados?*”. Entendemos que o Elemento Versionamento deva entrar nesta fase, visto que cada refinamento gera uma nova versão do conjunto de dados. Quanto ao questionamento sobre um novo Elemento Limpeza, não há como realizar.

Figura 37 – Avaliação dos Objetivos da Dimensão Refinamento.



Fonte: A autora (2019)

Figura 38 – Avaliação dos Níveis da Dimensão Refinamento.



Fonte: A autora (2019)

PA5 selecionou “Não, um ou mais elementos precisam ser incluídos” e comentou “*Enriquecer. Segurança do conjunto de dados está na descrição. Se refere a que?*”. Replicamos a mesma respostas dos comentários anteriores, para o “Enriquecer”. Quanto ao termo “Segurança” na descrição da Dimensão, percebemos que foi um erro de digitação.

Quanto aos objetivos, com exceção do *PA4*, todos os participantes selecionaram “Sim, os objetivos estão claros e completos” e não foi feito nenhum comentário adicional (ver Figura 37). Já o participante *PA4* selecionou “Não, um ou mais objetivos precisam ser alterados” e comentou “*Já que o objetivo desta dimensão é o refinamento em si, por que colocar versionamento? Uma coisa está relacionada a outra, mas ambas não são refinamento...*”. Discordamos do ponto de vista do participante, pois seja por meio de

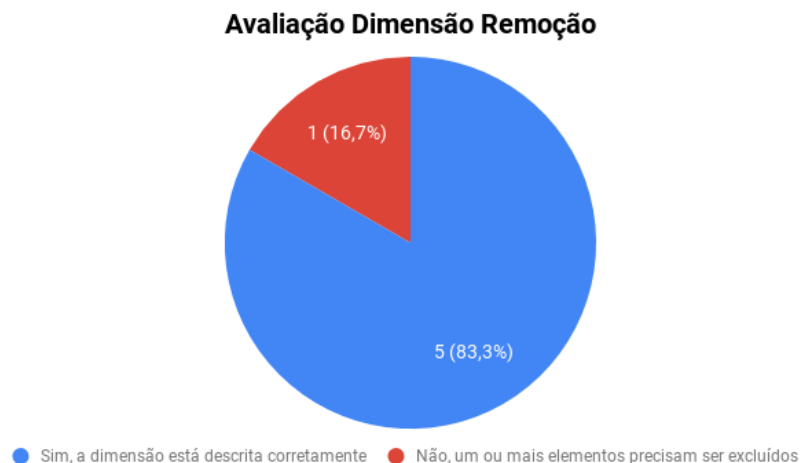
atualizações programadas ou não programadas, é gerada uma nova versão do conjunto de dados a partir do refinamento.

Por fim, quanto aos níveis (ver Figura 38), os participantes *PA1*, *PA2*, *PA3* e *PA6* selecionaram “Sim, os níveis estão definidos de forma clara e podem ser alcançados”. O *PA4* selecionou “Não, um ou mais níveis precisam ser alterados” e pontuou as mesmas coisas da resposta anterior. O participante *PA5* selecionou “Não, um ou mais níveis precisam ser incluídos” e comentou “*Acho que vale a pena pontuar algo no nível 3 de enriquecimento*”. Esse comentário foi atendido.

5.2.3.8 Avaliação da Dimensão Remoção

Para finalizar o conjunto de perguntas em relação às Dimensões do DWMM, questionamos novamente sobre a descrição, os objetivos e níveis da Dimensão Remoção. Mais uma vez, pretendíamos identificar lacunas, erros e/ou possibilidades de melhorias.

Figura 39 – Avaliação da Dimensão Remoção.

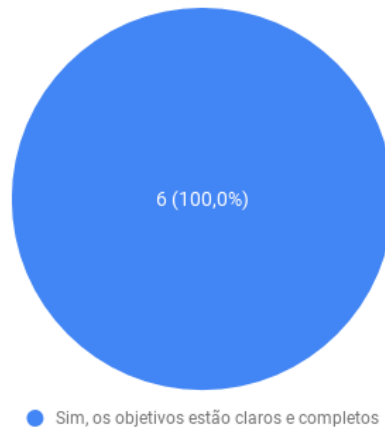


Fonte: A autora (2019)

Quanto à descrição da Dimensão (ver gráfico na Figura 39) os participantes *PA1*, *PA2*, *PA3*, *PA5* e *PA6* selecionaram “Sim, a dimensão está descrita corretamente” e não fizeram nenhum comentário adicional. O *PA4* selecionou “Não, um ou mais elementos precisam ser excluídos” e comentou “*Senti falta de desafios relacionados a viabilização da remoção e da persistência dos dados*”. Acreditamos que ele queria selecionar a opção “Não, um ou mais elementos precisam ser incluídos”, entretanto, este comentário já é atendido por meio do desafio de Preservação. Desta forma, não há alteração a ser feita.

Quanto aos objetivos da Dimensão, unanimemente, os participantes selecionaram “Sim, os objetivos estão claros e completos” e não fizeram nenhum comentário adicional (ver Figura 40).

Figura 40 – Avaliação dos Objetivos da Dimensão Remoção.

Avaliação dos Objetivos da Dimensão Remoção

Fonte: A autora (2019)

Figura 41 – Avaliação dos Níveis da Dimensão Remoção.

Avaliação dos Níveis da Dimensão Remoção

Fonte: A autora (2019)

Por fim, com relação aos níveis apenas o participante *PA4* selecionou “Não, um ou mais níveis precisam ser alterados” (ver Figura 41) e comentou “*Replico a resposta anterior para a mesma pergunta. De uma forma geral está mais voltado para como fazer do que o que fazer. Além disso, esta dimensão está apresentando mais ‘como’ deve ser alcançado para cada nível do que ‘o que’ deveria ser feito*”. Segundo (BECKER et al., 2010) há modelos de maturidade que podem assumir tanto natureza descritiva como prescritiva. Desta forma, decidimos continuar com esta visão.

Tabela 31 – Perguntas Gerais sobre o DWMM

	DWMM é útil para avaliar a maturidade dos conjuntos de dados na Web	DWMM é útil para avaliar o processo de publicação de dados na Web
PA1	5	5
PA2	5	5
PA3	5	3
PA4	3	4
PA5	4	4
PA6	5	5
Média	4,5	4,33

Fonte: A autora (2019)

5.2.3.9 Perguntas Gerais

A última fase do Grupo Focal foi composta por duas perguntas discursivas. Numa escala de 1 a 5, na qual 1 significa “Discordo Fortemente” e 5 significa “Concordo Fortemente”, solicitamos aos participantes que avaliassem as seguintes afirmações:

1. O DWMM é útil para avaliar a maturidade dos conjuntos de dados na Web?
2. O DWMM é útil para avaliar o processo de publicação de dados na Web?

A resposta de cada participante a essas perguntas é apresentada na Tabela 31.

Os participantes *PA1* e *PA2* não fizeram comentários em relação as perguntas. *PA3* comentou “*Em geral o trabalho está muito bom. Precisa de um pouco mais de atenção na padronização da descrição dos níveis...*”.

PA4 comentou “Como pontuei em várias respostas, há uma confusão em relação ao foco do trabalho. Acho que também o conjunto de dimensões acabou por prejudicar a forma de implantação e avaliação dos mesmos. Em relação a última pergunta, como ela está muito genérica, acredito que ele pode ser útil. Mas, acho que não é a melhor solução para avaliação. Já a primeira pergunta, tenho uma resposta neutra, pois acho que o modelo de maturidade de forma geral está considerando como grão 1 conjunto de dados”.

PA5 comentou “Qual a unidade de análise da avaliação? Processo/ciclo de vida? Maturidade do processo ou do conjunto de dados? O que seria maturidade do conjunto de dados?”.

PA6 comentou “Manter um padrão de escrita na descrição dos níveis; Revisar os objetivos do DWMM: Avaliar os conjuntos de dados ou o processo de uma organização publicar dados? Como esse modelo seria usado de fato para avaliar uma organização?”.

5.3 CONSIDERAÇÕES FINAIS

Esta pesquisa tinha como objetivo avaliar o DWLM a nível de viabilidade, completude e adequação. No geral, os participantes se mostraram satisfeitos com o modelo proposto. Alguns pontos de melhoria foram sugeridos, uns aceitos outros ficaram para implementação futura. Também reconhecemos as limitações do modelo, mas acreditamos que a escolha do DWLM, das Boas Práticas e dos Desafios para integrar os componentes do DWMM foi aceita de uma maneira geral. A nova versão do DWMM a partir das melhorias e sugestões aceitas após o Grupo Focal foi apresentada no Capítulo 4. Vale salientar que houve uma dúvida quanto ao objetivo do DWMM que, inicialmente, eram dois. Analisando os comentários preenchidos nos formulários e a partir da análise do áudio captado, reconsideramos e definimos como objetivo do DWMM a avaliação de organizações ou indivíduos no processo de publicação de dados na Web.

Embora grande parte das sugestões tenham sido incluídas, algumas foram registradas para serem consideradas posteriormente e outras julgamos como não necessárias.

6 CONCLUSÃO

Neste capítulo descrevemos as considerações finais sobre esta dissertação. Apresentamos as contribuições de pesquisas alcançadas (Seção 6.1) e as diretrizes de trabalhos futuros (Seção 6.2).

6.1 CONSIDERAÇÕES FINAIS

Neste trabalho, propomos um Modelo de Maturidade para Dados na Web. O modelo proposto se mostra relevante para o contexto de publicação e consumo de dados na Web, pois apresenta uma estrutura para guiar e apoiar organizações e indivíduos no processo de publicação de dados na Web. Além disso, descreve as Dimensões, que são as fases do Ciclo de Vida dos Dados na Web, e seus Elementos que são os desafios que surgem ao se publicar dados na Web. O modelo incorpora as boas práticas para Dados na Web em todas as Dimensões a fim de auxiliar organizações e indivíduos a enfrentar os desafios dos Dados na Web.

Inicialmente, foi apresentada uma visão geral dos Dados na Web, definindo conceitos que estão incorporados a esse termo mais geral, como Dados Abertos e Dados Conectados. Em seguida apresentamos as boas práticas para Dados na Web e como elas podem ser utilizadas para gerar benefícios tanto para os provedores quanto para os consumidores de dados, além de serem fundamentais para enfrentar os desafios que surgem ao se publicar e consumir dados na Web. Também foi apresentado uma visão geral do Modelo de Ciclo de Vida para Dados na Web (*Data on the Web Lifecycle Model*) - DWLM que é um dos principais elementos que compõe a arquitetura do DWMM.

No Capítulo 3 dissertamos sobre o que é a maturidade de uma maneira geral e o que constituem modelos de maturidade. Além disso, também apresentamos alguns modelos de maturidade tanto na área de Software quanto na área de Dados Abertos. No entanto, vimos que os modelos discutidos apresentam enfoques mais governamentais do que técnicos e possuem uma visão mais organizacional do processo. Juntamente com esta análise, o mapeamento sistemático realizado em (SANTOS et al., 2018), nos ajudou a identificar uma lacuna na área de modelos de maturidade para dados na Web que tratasse o processo de publicação de dados do ponto de vista técnico.

Desse modo, propomos um Modelo de Maturidade para Dados na Web que tem como objetivo avaliar e auxiliar organizações e indivíduos no processo de publicação de dados na Web. De uma forma descritiva e prescritiva, ele fornece um caminho evolutivo para se atingir a maturidade indicando como as boas práticas podem ser aplicadas para enfrentar os diferentes desafios ao se publicar dados na Web. O modelo proposto, intitulado DWMM, é composto por 3 níveis de maturidade, 6 dimensões que possuem de 1 a 6 elementos que

precisam ser avaliados de forma conjunta para se atingir a maturidade em determinado nível.

Para avaliarmos o DWMM realizamos um grupo focal que contou com 6 participantes da área de Dados na Web e de Modelos de Maturidade que tinham experiência na publicação e consumo de dados. Esse grupo focal foi realizado no Centro de Informática da Universidade Federal de Pernambuco e sua condução foi dividida em três etapas. Na primeira etapa foi solicitado que os participantes respondessem ao questionário a respeito dos seus dados acadêmicos e profissionais. A segunda etapa compreendeu a aplicação de um questionário contendo 22 perguntas a fim de avaliar as dimensões, os elementos, os níveis e os objetivos de cada dimensão do DWMM. Na terceira e última etapa foram feitas duas perguntas gerais com o intuito de avaliar se o modelo proposto era útil.

Por meio da avaliação, coletamos evidências sobre a importância do Modelo de Maturidade para Dados na Web, além de recebermos pontos de melhorias e um bom *feedback*. Após essa avaliação o modelo foi refinado e algumas das melhorias sugeridas pelos participantes foram incorporadas.

Entendemos que o DWMM possui limitações devido à escolha do DWLM e das boas práticas, mas acreditamos que, para sua proposta, é um modelo simples de ser aplicado e pode auxiliar as organizações e indivíduos a produzirem conjuntos de dados úteis na Web fornecendo uma visão de seu status e oferecendo meios que possibilitam a conquista da maturidade.

Vale salientar que o seguinte artigo foi publicado durante o Mestrado e nos auxiliou na identificação dessa lacuna de pesquisa, o que resultou na proposta deste trabalho:

- Santos, H.D.A.; Oliveira, M.I.S.; Lima, G.F.A.B.; Silva, K.M.; Muniz, R.I.V.C.S.; Lóscio, B.F. (2018) **Investigations into data published and consumed on the Web: A Systematic Mapping Study**. Journal of the Brazilian Computer Society, v. 24, n. 1, p. 14.

6.2 TRABALHOS FUTUROS

Como trabalhos futuros, identificamos as seguintes questões:

- **Desenvolver o Método de Avaliação do DWMM:** O método de avaliação descreve o processo de avaliação e os requisitos para atender ao DWMM. Desta forma, para ter uma melhor visão da importância do modelo proposto e fornecer maior clareza à organização e/ou indivíduo que utilizar o DWMM, é importante ter um modelo de autoavaliação que determine o nível atual de maturidade e que defina metas simples e objetivas para auxiliar o usuário a alcançá-las. Além de fornecer um monitoramento do progresso evolutivo da organização.

- **Revisar e/ou estabelecer novos níveis de maturidade:** Alguns dos comentários feitos durante o grupo focal nos estimulou a evoluir o DWMM para possuir mais um nível com atividades otimizadas.
- **Elaborar um portal para disponibilizar toda a documentação do DWMM:** Como já mencionado, a criação de um portal é de suma importância para criar um canal de comunicação para troca de experiências nos interessados do DWMM, além de contribuir com *feedback* que nos ajude a melhorar o trabalho.
- **Replicar a avaliação:** A replicação do grupo focal com outros grupos distintos de participantes e/ou a realização de um *survey* permitirá reunir novas evidências tanto para realização de melhorias no modelo, quanto para a evolução do mesmo.
- **Realizar estudos de caso:** A realização de estudos de caso pode nos ajudar a ter uma melhor visão da importância do modelo proposto e observar, na prática, o comportamento e desempenho do modelo durante sua aplicação.

REFERÊNCIAS

- ABITEBOUL S., B. P.; SUCIU, D. *Data on the Web: from relations to semistructured data and XML*. [S.l.]: Morgan Kaufmann, 2000.
- ALSHAMMARI, M.; SIMPSON, A. J. R. Personal data management for privacy engineering: An abstract personal data lifecycle model. 2017.
- ANDERSON, E. S.; JESSEN, S. A. Project maturity in organizations. *International Journal of Project Management*, v. 21, n. 6, p. 457–461, 2003.
- BACKES, D. S.; COLOMÉ, J. S.; ERDMANN, R. H.; LUNARDI, V. L. Grupo focal como técnica de coleta e análise de dados em pesquisas qualitativas. *O mundo da saúde*, São Paulo, v. 35, n. 4, p. 438–42, 2011.
- BARNAGHI, P.; SHETH, A.; HENSON, C. From data to actionable knowledge: big data challenges in the web of things. *IEEE Intelligent Systems*, IEEE, n. 6, p. 6–11, 2013.
- BECKER, J.; KNACKSTEDT, R.; PÖPPELBUSS, J. Developing maturity models for it management. *Business & Information Systems Engineering*, Springer, v. 1, n. 3, p. 213–222, 2009.
- BECKER, J.; NIEHAVES, B.; POEPELBUSS, J.; SIMONS, A. Maturity models in is research. 2010.
- BERNERS-LEE, T. *Linked Data*. 2006. <<https://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em 26 de novembro de 2018.
- BERNERS-LEE T., C. D.; SWICK, R. R. *Web architecture: Describing and exchanging data*. 1999. Disponível em: <<https://www.w3.org/1999/04/WebData>>. Acesso em: 27 mai. 2019.
- BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked data-the story so far. *Semantic services, interoperability and web applications: emerging concepts*, p. 205–227, 2009.
- BOUQUET, P.; STOERMER, H. Web of data and web of entities: Identity and reference in interlinked data in the semantic web. *Philosophy & Technology*, v. 25, p. 5–26, 2012.
- BRUIN, T. D.; FREEZE, R.; KAULKARNI, U.; ROSEMANN, M. Understanding the main phases of developing a maturity assessment model. Australasian Chapter of the Association for Information Systems, 2005.
- CAPLAN, S. Using focus group methodology for ergonomic design. *Ergonomics*, Taylor & Francis, v. 33, n. 5, p. 527–533, 1990.
- CARALLI, R.; KNIGHT, M.; MONTGOMERY, A. *Maturity models 101: A primer for applying maturity models to smart grid security, resilience, and interoperability*. [S.l.], 2012.
- CATTEAU, O.; VIDAL, P.; BROISIN, J. A generic representation allowing for expression of learning object and metadata lifecycle. In: *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*. Kerkrade, Netherlands: [s.n.], 2006. p. 30–32. ISSN 2161-3761.

- CHEN, Y.; CHEN, S.; LIN, S. C. A metadata lifecycle model for digital libraries : methodology and application for an evidence-based approach to library research. In: *World Library and Information Congress: 69th IFLA General Conference and Council*. [S.l.: s.n.], 2003.
- CHIARA, I. G. D. Grupo de foco. In: *Métodos qualitativos de pesquisa em Ciência da Informação*. São Paulo, Brasil: Polis, 2005. p. 101–117. ISBN 85-7228-021-9.
- DERILINX, D. L.; LÓSCIO, B. F.; ARCHER, P. W3C Working Group Note, *Data on the Web Best Practices Use Cases & Requirements*. 2015. <<https://www.w3.org/TR/dwbp-ucr/>>. Acesso em 18 de dezembro de 2018.
- DIETRICH, D.; GRAY, J.; MCNAMARA, T.; POIKOLA, A.; POLLOCK, R.; TAIT, J.; ZIJLSTRA, T. *Open Data Handbook*. [S.l.: Open Knowledge Foundation, 2009. <<http://opendatahandbook.org/guide/en/>>. Acesso em 26 de novembro de 2018.
- DODDS, L.; NEWMAN, A. *A guide to the Open Data Maturity Model - Assessing your open data publishing and use*. [S.l.: Open Data Institute, 2015. <<https://theodi.org/article/open-data-maturity-model-2/>>. Acesso em 03 de abril de 2019.
- FISHER, D. M. The business process maturity model: a practical approach for identifying opportunities for optimization. *Business Process Trends*, v. 9, n. 4, p. 11–15, 2004.
- FITTERER, R.; ROHNER, P. Towards assessing the networkability of health care providers: a maturity model approach. *Information Systems and E-Business Management*, Springer, v. 8, n. 3, p. 309–333, 2010.
- FRASER, P.; MOULTRIE, J.; GREGORY, M. The use of maturity models/grids as a tool in assessing product development capability. In: IEEE. *IEEE international engineering management conference*. [S.l.], 2002. v. 1, p. 244–249.
- GRAY, J.; CHAMBERS, L.; BOUNEGRU, L. *The data journalism handbook: how journalists can use data to improve the news*. [S.l.: "O'Reilly Media, Inc.", 2012.
- HALPER, F.; KRISHNAN, K. Tdwi big data maturity model guide. *Interpreting Your Assessment Score. TDWI Research*, v. 1, p. 16, 2013.
- HEVNER, A. R.; MARCH, S. T.; PARK, J.; RAM, S. Design science in information systems research. *MIS quarterly*, Springer, v. 28, n. 1, p. 75–105, 2004.
- HÜNER, K. M.; OFNER, M.; OTTO, B. Towards a maturity model for corporate data quality management. In: ACM. *Proceedings of the 2009 ACM symposium on Applied Computing*. [S.l.], 2009. p. 231–238.
- ISOTANI, S.; BITTENCOURT, I. I. *Dados Abertos Conectados: Em busca da Web do Conhecimento*. [S.l.: Novatec Editora, 2015. Disponível em: <<https://ceweb.br/livros/dados-abertos-conectados/>>. Acesso em: 02 jul. 2019.
- JOSHI, P.; ISLAM, S. E-government maturity model for sustainable e-government services from the perspective of developing countries. *Sustainability*, Multidisciplinary Digital Publishing Institute, v. 10, n. 6, p. 1882, 2018.
- KRUEGER, R. A. *Focus groups: A practical guide for applied research*. 2. ed. [S.l.: Sage publications, 1994.

- KRUEGER, R. A. *Focus groups: A practical guide for applied research*. [S.l.]: Sage Publications, 2014.
- LASRADO, L. A.; VATRAPU, R.; ANDERSEN, K. N. Maturity models development in is research: a literature review. In: *IRIS Selected Papers of the Information Systems Research Seminar in Scandinavia 2015. Paper*. [S.l.: s.n.], 2015. v. 6.
- LEE, G.; KWAK, Y. H. An open government maturity model for social media-based public engagement. *Government information quarterly*, Elsevier, v. 29, n. 4, p. 492–503, 2012.
- LÓSCIO, B.; GUIMARÃES, C.; CALEGARI, N. Data on the web best practices: Challenges and benefits. *Open Data Reserach Symposium (ODRS 2016)*, 2016.
- LÓSCIO, B. F.; BURLE, C.; CALEGARI, N. W3C Recommendation, *Data on the Web Best Practices*. 2016. <<https://www.w3.org/TR/dwbp/>>. Acessado em 15 de dezembro de 2018.
- LÓSCIO, B. F.; BURLE, C.; IURY, M. S.; CALEGARI, N. C. *Fundamentos para publicação de dados na Web*. [S.l.]: Comitê Gestor da Internet no Brasil CGI.br, 2018. ISBN 978-85-5559-072-6.
- LÓSCIO, B. F.; GUIMARÃES, C. B. S.; CALEGARI, N. J. Boas práticas para dados na web: desafios e benefícios. *Revista Principia*, 2016.
- LÓSCIO, B. F.; OLIVEIRA, M. I. S.; BITTENCOURT, I. I. Publicação e Consumo de Dados na Web: Conceitos e Desafios. *Tópicos em Gerenciamento de Dados e Informações (Mini Cursos - SBBD 2015)*, d, p. 39–69, 2015. Disponível em: <<http://dexl.incc.br/sbbd2015/anais/ShortCourses.pdf>>.
- MADHAVAN, J.; JEFFERY, S. R.; COHEN, S.; DONG, X. L.; KO, D.; YU, C.; HALEVY, A. Web-scale data integration: You can only afford to pay as you go. 2007.
- MANUAL, D. D. A. Manual dos dados abertos: desenvolvedores. *Cooperação técnica científica entre Laboratório Brasileiro de Cultura Digital e o Núcleo de Informação e Coordenação do Ponto BR (NIC. br)*. São Paulo: Comitê Gestor da Internet no Brasil, 2011.
- METTLER, T. A design science research perspective on maturity models in information systems. Institute of Information Management, Universtiy of St. Gallen, 2009.
- METTLER, T.; ROHNER, P. Situational maturity models as instrumental artifacts for organizational design. In: ACM. *Proceedings of the 4th international conference on design science research in information systems and technology*. [S.l.], 2009. p. 22.
- METTLER, T.; ROHNER, P.; WINTER, R. Towards a classification of maturity models in information systems. In: *Management of the interconnected world*. [S.l.]: Springer, 2010. p. 333–340.
- MORGAN, D. L. *Focus groups as qualitative research*. [S.l.]: Sage publications, 1996. v. 16.
- MÖLLER, K. Lifecycle models of data-centric systems and domains. *Semantic Web*, v. 4, p. 67–88, 2013. Disponível em: <<http://doi.org/10.3233/SW.2012.0060>>.

- NECASKÝ, M.; CHLAPEK, D.; KLÍMEK, J.; KUCERA, J.; MAURINO, A.; RULA, A.; KONECNY, M.; VANOVA, L. Methodology for publishing datasets as open data. *DELIVERABLE D5.1*, 2013. Disponível em: <https://www.comsode.eu/wp-content/uploads/D5.1-Methodology_for_publishing_datasets_as_open_data.pdf>.
- NETO, H.; MAGALHÃES, E.; MOURA, H.; CAPPELLI, C.; MARTINS, L. et al. Avaliação de um modelo de maturidade para governança ágil em tic usando focus group. In: SBC. *Anais do XI Simpósio Brasileiro de Sistemas de Informação*. [S.l.], 2015. p. 15–22.
- NETO, H. R. d. A. *Um Modelo de Maturidade para Governança Ágil em Tecnologia da Informação e Comunicação*. Tese (Doutorado) — Universidade Federal de Pernambuco, Centro de Informática, Curso de Pós-Graduação em Ciências da Computação, Recife, 2015.
- NOLAN, R. L. Managing the computer resource: a stage hypothesis. *Communications of the ACM*, ACM, v. 16, n. 7, p. 399–405, 1973.
- OLIVEIRA, L. E. R.; OLIVEIRA, M. I. S.; SANTOS, W. C. d. R.; LÓSCIO, B. F. Data on the web management system: a reference model. In: ACM. *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. [S.l.], 2018. p. 2.
- OLIVEIRA, L. E. R. de A.; OLIVEIRA, M. I. S.; LÓSCIO, B. F. Um survey sobre soluções para publicação de dados na web sob a perspectiva das boas práticas do w3c. In: *32nd Brazilian Symposium on Databases*. [S.l.: s.n.], 2017.
- OLIVEIRA, M. I. S.; OLIVEIRA, H. R. de; OLIVEIRA, L. A.; LÓSCIO, B. F. Open government data portals analysis: the brazilian case. In: ACM. *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*. [S.l.], 2016. p. 415–424.
- PÁDUA, W. d. Engenharia de software: fundamentos, métodos e padrões. *Rio de Janeiro: LTC*, 2003.
- PAULK, M. C.; CURTIS, B.; CHRISSIS, M. B.; WEBER, C. V. *Capability Maturity Model for Software, Version 1.1*. Software Engineering Institute. [S.l.], 1993.
- PEFFERS, K.; TUUNANEN, T.; ROTHENBERGER, M. A.; CHATTERJEE, S. A design science research methodology for information systems research. *Journal of management information systems*, Taylor & Francis, v. 24, n. 3, p. 45–77, 2007.
- PIRES, M. T. Guia de dados abertos. *São Paulo: Este Guia é parte integrante do Projeto de Cooperação entre o Governo do Estado de São Paulo e o Reino Unido*, Comitê Gestor da Internet no Brasil CGI.br, 2015. Disponível em: <https://ceweb.br/media/docs/publicacoes/13/Guia_Dados_Abertos.pdf>. Acesso em: 02 jul. 2019.
- PROENÇA, D.; BORBINHA, J. Maturity models for data and information management. In: SPRINGER. *International Conference on Theory and Practice of Digital Libraries*. [S.l.], 2018. p. 81–93.

- RADULOVIC, F.; POVEDA-VILLALÓN, M.; VILA-SUERO, D.; RODRÍGUEZ-DONCEL, V.; GARCÍA-CASTRO, R.; GÓMEZ-PÉREZ, A. Guidelines for linked data generation and publication: An example in building energy consumption. *Automation in Construction*, v. 57, p. 178 – 187, 2015. ISSN 0926-5805. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0926580515000801>>.
- SAGRADA, B. Almeida corrigida fiel (acf). *SP: Sociedade Bíblica do*, 1994.
- SANTOS, H. *Uma Estratégia para o Refinamento de Dados na Web Baseada em Social Coding*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, Centro de Informática, Curso de Pós-Graduação em Ciências da Computação, Recife, 2018.
- SANTOS, H. D. A. d.; OLIVEIRA, M. I. S.; LIMA, G. d. F. A. B.; SILVA, K. M. da; MUNIZ, R. I. V. C. S.; LÓSCIO, B. F. Investigations into data published and consumed on the web: A systematic mapping study. *Journal of the Brazilian Computer Society*, v. 24, n. 1, p. 14, 2018. Disponível em: <<https://doi.org/10.1186/s13173-018-0077-z>>.
- SEI, S. E. I. Cmmi for services, version 1.3. software engineering institute. *Carnegie Mellon University, Pittsburg. Cooper, J and M Fisher (2002). Software Acquisition Capability Maturity Model, Version*, v. 1, p. 25–34, 2010.
- SEN, A.; RAMAMURTHY, K.; SINHA, A. P. A model of data warehousing process maturity. *IEEE Transactions on Software Engineering*, IEEE, v. 38, n. 2, p. 336–353, 2011.
- SILVA, E. C. G.; OLIVEIRA, M. I. S.; OLIVEIRA, E.; GAMA, K. S. da; LÓSCIO, B. F. Um survey sobre plataformas de mediação de dados para internet das coisas. *42º Seminário Integrado de Software e Hardware*, 2015.
- SILVA, K. M. da. *Um Modelo de Ciclo de Vida dos Dados na Web*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, Centro de Informática, Curso de Pós-Graduação em Ciências da Computação, Recife, 2019.
- SIMON, H. A. *The sciences of the artificial*. [S.l.]: MIT press, 1996.
- STERNKOPF, H.; MUELLER, R. M. Doing good with data: Development of a maturity model for data literacy in non-governmental organizations. In: *Proceedings of the 51st Hawaii International Conference on System Sciences*. [S.l.: s.n.], 2018.
- SULAIMAN, H.; COB, Z. C.; ALI, N. Big data maturity model for malaysian zakat institutions to embark on big data initiatives. In: IEEE. *2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*. [S.l.], 2015. p. 61–66.
- TREMBLAY, M. C.; HEVNER, A. R.; BERNDT, D. J. Focus groups for artifact refinement and evaluation in design research. *Cais*, v. 26, n. 27, p. 599–618, 2010.
- VASCONCELOS, A.; MORAIS, L. Modelos de maturidade para processos de software: Cmmi e mps. br. Centro de Informática - UFPE, 2010.
- VESSET, D.; GIRARD, G.; FEBLOWITZ, J.; VERSACE, M.; BURGHARD, C.; O'BRIEN, A.; OLOFSON, C.; SCHUBMEHL, D.; MCDONOUGH, B.; WOODWARD, A. et al. Idc maturityscape: Big data and analytics 2.0. *International Data Corporation*, 2015.

WE ARE SOCIAL. *DIGITAL 2019: GLOBAL DIGITAL OVERVIEW*. 2019. Disponível em: <<https://datareportal.com/reports/digital-2019-global-digital-overview>>. Acesso em: 20 jul. 2019.

WEBER, C.; CURTIS, B.; GARDINER, T. Business process maturity model (bpmm) version 1.0 <http://www.omg.org/docs/formal/08-06-01.pdf>. *Datum des Zugriffs*, p. 07–21, 2008.

WENDLER, R. The maturity of maturity model research: A systematic mapping study. *Information and software technology*, Elsevier, v. 54, n. 12, p. 1317–1339, 2012.

WILKINSON, S.; SILVERMAN, D. Focus group research. *Qualitative research: Theory, method and practice*, Sage, p. 177–199, 2004.

ZAGANELLI, B. M.; NISENBAUM, M. A.; MARQUES, S. B.; OLINTO, G. O grupo focal na ciência da informação. *Informação & Sociedade: Estudos*, p. 37–47, 2015.

APÊNDICE A – FORMULÁRIO DE AVALIAÇÃO DO DWMM

Meu nome é Rayelle Ingrid Vera Cruz Silva Muniz. Sou mestranda em Ciência da Computação no Centro de Informática (CIn) da Universidade Federal de Pernambuco (UFPE). Eu sou orientanda da professora Bernadette Farias Lóscio. Gostaria de agradecer por colaborar com o meu trabalho participando deste grupo focal. Seu feedback é extremamente valioso para a conclusão da minha pesquisa de mestrado.

O questionário a seguir avalia um Modelo de Maturidade para Dados na Web (Data on the Web Maturity Model - DWMM). Em particular, o modelo proposto é composto por um conjunto de dimensões, elementos para cada dimensão, níveis de maturidade, os quais foram pensados com o intuito de avaliar a maturidade dos provedores de dados em produzir conjuntos de dados úteis, bem como avaliar a maturidade dos conjuntos de dados gerados ao longo do seu ciclo de vida na Web.

Suas respostas são anônimas, ou seja, você não precisa fornecer suas informações pessoais, além do seu e-mail, se quiser receber os resultados dessa pesquisa.

FASE 1 - Informações sobre o Respondente

As informações abaixo servem meramente para **caracterização do perfil dos respondentes**. Em **nenhum momento os respondentes serão identificados ou relacionados individualmente dentro do trabalho**.

1. Qual o seu cargo/função atualmente?

2. Qual o seu nível de formação (concluído)?

- () Ensino Fundamental - 1º Grau
- () Ensino Médio - 2º Grau
- () Ensino Profissional - Técnico
- () Graduação
- () Especialização
- () Mestrado Acadêmico/Profissional
- () Doutorado
- () Pós-doutorado

3. Você tem participado direta ou indiretamente de atividades relacionadas a Dados na Web (Publicação e Consumo)?

☐ Não

☐ Sim

4. Em caso de sim, há quantos anos está participando de atividades relacionadas a Dados na Web (Publicação e Consumo)?

5. Você tem participado ou já participou direta ou indiretamente de atividades relacionadas a Modelos de Maturidade (Utilização ou Criação de um modelo)?

☐ Não

☐ Sim

6. Em caso de sim, há quantos anos está participando de atividades relacionadas a Modelos de Maturidade (Utilização ou Criação de um modelo)?

7. Você desenvolve suas atividades em qual meio?

☐ Acadêmico

☐ Profissional

☐ Ambos

FASE 2 - Instruções para preenchimento do questionário

O Modelo de Maturidade para Dados na Web denominado DWMM contém 6 Dimensões, entre 1 e 7 elementos para cada dimensão e 3 níveis de maturidade. Este questionário pretende apoiar a avaliação do modelo proposto por meio da realização de um grupo focal.

Para responder o presente questionário você deverá levar em consideração sua experiência na área de Publicação e Consumo de Dados na Web, bem como um conhecimento básico de um Modelo de Maturidade.

Níveis de Maturidade

Níveis de Maturidade são considerados um conjunto de características distintas (práticas, as medidas ou ações por dimensão) que são empiricamente testáveis.

1. Os níveis de maturidade proposto pelo DWMM são fáceis de serem interpretados?

A ideia aqui é identificar lacunas, erros e/ou possibilidades de melhoria.

- () Não, um nível não ficou claro
- () Não, todos os níveis estão complexos
- () Sim, os níveis estão descritos corretamente e é de fácil interpretação

Quais são as suas sugestões em caso de necessidade de falta de clareza ou complexidade?

2. O número de níveis é suficiente?

A ideia aqui é identificar se os níveis são viáveis, completos e adequados para o DWMM. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- () Não, um ou mais níveis precisam ser incluídos
- () Não, um ou mais níveis precisam ser excluídos
- () Não, um ou mais níveis precisam ser alterados
- () Sim, o número de níveis é suficiente

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão ou atualização?

3. Os nomes dos níveis são adequados para o DWMM?

A ideia aqui é identificar possibilidades de melhoria.

- () Não, um ou mais níveis precisam ser alterados
- () Sim, os nomes são adequados para o DWMM

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão ou atualização?

Dimensões DWMM

Dimensões são um conjunto de práticas que agrupam atributos semelhantes em uma

área que, quando executadas coletivamente, satisfazem a um conjunto de metas consideradas importantes para realização de melhorias significativas.

1. O conjunto de dimensões do DWMM está descrito corretamente?

Gostaríamos de identificar se as dimensões são viáveis, completas e adequadas para o DWMM. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, uma dimensão precisa ser excluída
- ☐ Não, uma dimensão precisa ser atualizada
- ☐ Não, uma dimensão precisa ser incluída
- ☐ Sim, o conjunto de dimensões está descrito corretamente

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão ou atualização?

Dimensão: Planejamento

1. A dimensão está descrita corretamente?

Gostaríamos de identificar se os elementos desta dimensão são viáveis, completos e adequados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais elementos precisam ser excluídos
- ☐ Não, um ou mais elementos precisam ser atualizados
- ☐ Não, um ou mais elementos precisam ser incluídos
- ☐ Sim, a dimensão está descrita corretamente

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

2. Os objetivos da dimensão estão claros?

A ideia aqui é identificar se os objetivos são claramente definidos para cada nível, se eles são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais objetivos precisam ser alterados

- ☐ Não, um ou mais objetivos precisam ser excluídos
- ☐ Não, um ou mais objetivos precisam ser incluídos
- ☐ Sim, os objetivos estão claros e completos

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

3. Os níveis estão claramente definidos?

A ideia aqui é identificar se os níveis são descritos de forma clara e concisa, se são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais níveis precisam ser alterados
- ☐ Não, um ou mais níveis precisam ser incluídos
- ☐ Não, um ou mais níveis precisam ser excluídos
- ☐ Sim, os níveis estão definidos de forma clara e podem ser alcançados

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

Dimensão: Criação

1. A dimensão está descrita corretamente?

Gostaríamos de identificar se os elementos desta dimensão são viáveis, completos e adequados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais elementos precisam ser excluídos
- ☐ Não, um ou mais elementos precisam ser atualizados
- ☐ Não, um ou mais elementos precisam ser incluídos
- ☐ Sim, a dimensão está descrita corretamente

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

2. Os objetivos da dimensão estão claros?

A ideia aqui é identificar se os objetivos são claramente definidos para cada nível, se eles são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais objetivos precisam ser alterados
- ☐ Não, um ou mais objetivos precisam ser excluídos
- ☐ Não, um ou mais objetivos precisam ser incluídos
- ☐ Sim, os objetivos estão claros e completos

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

3. Os níveis estão claramente definidos?

A ideia aqui é identificar se os níveis são descritos de forma clara e concisa, se são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais níveis precisam ser alterados
- ☐ Não, um ou mais níveis precisam ser incluídos
- ☐ Não, um ou mais níveis precisam ser excluídos
- ☐ Sim, os níveis estão definidos de forma clara e podem ser alcançados

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

Dimensão: Publicação

1. A dimensão está descrita corretamente?

Gostaríamos de identificar se os elementos desta dimensão são viáveis, completos e adequados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais elementos precisam ser excluídos
- ☐ Não, um ou mais elementos precisam ser atualizados
- ☐ Não, um ou mais elementos precisam ser incluídos
- ☐ Sim, a dimensão está descrita corretamente

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

2. Os objetivos da dimensão estão claros?

A ideia aqui é identificar se os objetivos são claramente definidos para cada nível, se eles são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais objetivos precisam ser alterados
- ☐ Não, um ou mais objetivos precisam ser excluídos
- ☐ Não, um ou mais objetivos precisam ser incluídos
- ☐ Sim, os objetivos estão claros e completos

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

3. Os níveis estão claramente definidos?

A ideia aqui é identificar se os níveis são descritos de forma clara e concisa, se são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais níveis precisam ser alterados
- ☐ Não, um ou mais níveis precisam ser incluídos
- ☐ Não, um ou mais níveis precisam ser excluídos
- ☐ Sim, os níveis estão definidos de forma clara e podem ser alcançados

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

Dimensão: Consumo

1. A dimensão está descrita corretamente?

Gostaríamos de identificar se os elementos desta dimensão são viáveis, completos e adequados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de me-

lhoria.

- ☐ Não, um ou mais elementos precisam ser excluídos
- ☐ Não, um ou mais elementos precisam ser atualizados
- ☐ Não, um ou mais elementos precisam ser incluídos
- ☐ Sim, a dimensão está descrita corretamente

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

2. Os objetivos da dimensão estão claros?

A ideia aqui é identificar se os objetivos são claramente definidos para cada nível, se eles são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais objetivos precisam ser alterados
- ☐ Não, um ou mais objetivos precisam ser excluídos
- ☐ Não, um ou mais objetivos precisam ser incluídos
- ☐ Sim, os objetivos estão claros e completos

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

3. Os níveis estão claramente definidos?

A ideia aqui é identificar se os níveis são descritos de forma clara e concisa, se são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais níveis precisam ser alterados
- ☐ Não, um ou mais níveis precisam ser incluídos
- ☐ Não, um ou mais níveis precisam ser excluídos
- ☐ Sim, os níveis estão definidos de forma clara e podem ser alcançados

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

Dimensão: Refinamento

1. A dimensão está descrita corretamente?

Gostaríamos de identificar se os elementos desta dimensão são viáveis, completos e adequados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais elementos precisam ser excluídos
- ☐ Não, um ou mais elementos precisam ser atualizados
- ☐ Não, um ou mais elementos precisam ser incluídos
- ☐ Sim, a dimensão está descrita corretamente

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

2. Os objetivos da dimensão estão claros?

A ideia aqui é identificar se os objetivos são claramente definidos para cada nível, se eles são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais objetivos precisam ser alterados
- ☐ Não, um ou mais objetivos precisam ser excluídos
- ☐ Não, um ou mais objetivos precisam ser incluídos
- ☐ Sim, os objetivos estão claros e completos

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

3. Os níveis estão claramente definidos?

A ideia aqui é identificar se os níveis são descritos de forma clara e concisa, se são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais níveis precisam ser alterados
- ☐ Não, um ou mais níveis precisam ser incluídos
- ☐ Não, um ou mais níveis precisam ser excluídos
- ☐ Sim, os níveis estão definidos de forma clara e podem ser alcançados

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

Dimensão: Remoção

1. A dimensão está descrita corretamente?

Gostaríamos de identificar se os elementos desta dimensão são viáveis, completos e adequados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais elementos precisam ser excluídos
- ☐ Não, um ou mais elementos precisam ser atualizados
- ☐ Não, um ou mais elementos precisam ser incluídos
- ☐ Sim, a dimensão está descrita corretamente

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

2. Os objetivos da dimensão estão claros?

A ideia aqui é identificar se os objetivos são claramente definidos para cada nível, se eles são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- ☐ Não, um ou mais objetivos precisam ser alterados
- ☐ Não, um ou mais objetivos precisam ser excluídos
- ☐ Não, um ou mais objetivos precisam ser incluídos
- ☐ Sim, os objetivos estão claros e completos

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

3. Os níveis estão claramente definidos?

A ideia aqui é identificar se os níveis são descritos de forma clara e concisa, se são completos e podem ser alcançados. Também gostaríamos de identificar lacunas, erros

e/ou possibilidades de melhoria.

- () Não, um ou mais níveis precisam ser alterados
- () Não, um ou mais níveis precisam ser incluídos
- () Não, um ou mais níveis precisam ser excluídos
- () Sim, os níveis estão definidos de forma clara e podem ser alcançados

Quais são as suas sugestões em caso de necessidade de inclusão, exclusão, agrupamento ou atualização?

FASE 3 - Perguntas Discursivas

Numa escala de 1 a 5, na qual 1 significa “Discordo Fortemente” e 5 significa “Concordo Fortemente”, como você avalia as afirmações abaixo:

1. O DWMM é útil para avaliar a maturidade dos conjuntos de dados na Web

Gostaríamos de identificar se, para você, o DWMM consegue avaliar o nível de maturidade dos conjuntos de dados a partir da avaliação de como os desafios na Web são enfrentados em cada etapa do DWLM.

2. O DWMM é útil para avaliar o processo de publicação de dados na Web

Gostaríamos de identificar se, para você, o DWMM consegue avaliar a maturidade do processo de publicação de dados na Web.

Demais sugestões para o DWMM? Favor descrever abaixo:
