



Pós-Graduação em Ciência da Computação

Glória de Fátima Andrade Barros Lima

Um Framework para Avaliação da Saúde de Ecossistemas de Dados



Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
www.cin.ufpe.br/posgraduacao

RECIFE

2019

Glória de Fátima Andrade Barros Lima

Um Framework para Avaliação da Saúde de Ecossistemas de Dados

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
Coorientador: Prof. Dr. Marcelo Iury de Sousa Oliveira

Recife
2019

Catálogo na fonte
Bibliotecária Mariana de Souza Alves CRB4-2105

L732f Lima, Glória de Fátima Andrade Barros
Um Framework para Avaliação da Saúde de Ecossistemas de
Dados/ Glória de Fátima Andrade Barros Lima – 2019.
133 f., fig., tab.

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

1. Banco de Dados 2. Framework. 3. Avaliação da Saúde. 4.
Ecossistemas de Dados. I. Lóscio, Bernadette Farias
(orientadora). II. Título.

025.04

CDD (22. ed.)

UFPE-CCEN 2020-03

Glória de Fátima Andrade Barros Lima

“Um Framework para Avaliação da Saúde de Ecossistemas de Dados”

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: 22 de agosto de 2019.

Orientadora: Profa. Dra. Bernadette Farias Lóscio

BANCA EXAMINADORA

Profa. Dra. Carina Frota Alves
Centro de Informática / UFPE

Prof. Dr. José Viterbo Filho
Instituto de Computação/UFF

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

Dedico esta Dissertação a minha mãe e ao meu pai.

AGRADECIMENTOS

O caminho percorrido até chegar ao final foi povoado por desafios pessoais e profissionais, aprendizado, conquistas e principalmente, amadurecimento. Agradeço a Deus por ter me dado força e saúde para concluir mais essa etapa! Gostaria de agradecer a várias pessoas que contribuíram de diversas formas para a conclusão desta dissertação.

Em primeiro lugar, gostaria de agradecer à minha mãe por sempre me apoiar, ser meu exemplo de força e coragem, e sempre me estimular a alcançar os meus objetivos. Agradeço ao meu pai pelos ensinamentos, exemplos e amor que seguem comigo. Agradeço também a minha família por ser uma fonte de carinho e inspiração no meu dia-a-dia.

Agradeço ao meu amor Diogo, pelo seu carinho, dedicação e sabedoria em todos os momentos. Obrigada por ser quem você é e por sempre me apoiar e incentivar a encarar novos desafios. Sua ajuda e apoio nesses dois anos foram essenciais!

Gostaria de agradecer a minha orientadora, Prof^ª Bernadette Farias Lóscio (Berna), e ao meu co-orientador, Prof. Marcelo Iury Oliveira, por terem me orientado, me apoiado e confiado em mim durante toda essa jornada. Agradeço principalmente por terem acreditado em mim até nos momentos em que eu mesma não acreditava. Muito obrigada por tudo!

Agradeço aos amigos e integrantes do grupo de pesquisa aLADIN, Karina, Rayelle, Wilker, Lairson, Vitória, Ruan, Grennda e Helton, por terem me ajudado de forma direta ou indireta com revisões, discussões e orientações nessa dissertação, sempre se mostrando dispostos a ajudar. Karina e Rayelle, obrigada por sempre se mostrarem pacientes e disponíveis todas as vezes em que precisei, seja para tirar dúvidas ou, até mesmo, para desabafar sobre a ansiedade de chegar ao fim desse ciclo. Muito obrigada pessoal!

Agradeço aos meus colegas de trabalho pela compreensão durante os últimos meses. Pois, sempre que precisei me ausentar para a realização desta dissertação, tive o apoio e a compreensão de todos. Muito obrigada! Em especial, agradeço aos meus líderes pelo suporte e flexibilidade que me foram dados.

Por último, e não menos importante, gostaria de agradecer a todos que de alguma forma contribuíram para que esta dissertação fosse realizada. Muito obrigada!!

RESUMO

A disponibilização de dados em meio digital, por entidades públicas e privadas, tem crescido bastante nos últimos anos, contribuindo para a geração de valor por meio do compartilhamento e consumo desses dados. Nesse contexto, surgem os Ecossistemas de Dados (ED), que podem ser definidos como redes de atores autônomos que consomem, produzem ou fornecem direta ou indiretamente dados e outros recursos relacionados aos dados (e.g., software, serviços e infraestrutura). Apesar do grande número de Ecossistemas de Dados disponíveis atualmente, nem todos podem ser considerados saudáveis, dificultando a aquisição de novas parcerias, investimentos e impactando na produção de consumo de recursos pelos atores do ecossistema. Dessa forma, avaliar a saúde desses ecossistemas torna-se fundamental para prevenir problemas no seu funcionamento, bem como para garantir a sua sobrevivência ao longo do tempo. A saúde de um ED pode ser avaliada por meio dos indicadores de produtividade, robustez, criação de nicho e sustentabilidade, os quais revelam um panorama da saúde do ecossistema. Neste trabalho propomos um framework para avaliação da saúde de Ecossistemas de Dados (FASSED) composto por indicadores, os quais avaliam o estado atual e a capacidade de se manter saudável ao longo do tempo. Cada indicador utiliza métricas que permitem a avaliação dos principais elementos do ED e permite sua aplicação de forma contínua para avaliar a saúde do ecossistema em análise. Com o uso do framework proposto, espera-se obter indícios sobre o estado dos indicadores para auxiliar nas tomadas de decisão relacionadas aos processos de publicação e consumo de dados, bem como avaliar quais ecossistemas são mais prósperos ou carecem de maiores investimentos.

Palavras-chaves: Framework. Avaliação da Saúde. Ecossistemas de Dados.

ABSTRACT

The availability of data in digital media, by public and private entities, has greatly increased in recent years, contributing to value creation through data sharing and consumption. In this context, Data Ecosystems emerge, which can be defined as networks of autonomous actors that directly or indirectly consume, produce, or provide data and other data-based resources (*e.g.*, software, services, and infrastructure). Despite the large number of data ecosystems available, not all of them can be considered healthy, making it difficult to acquire new partnerships, investments and impacting the production of resource consumption by the ecosystem actors. Thus, assessing the ecosystem health becomes critical to prevent problems in their functioning as well as to ensure their survival over time. The health of a Data Ecosystem can be assessed by the indicators of productivity, robustness, niche creation and sustainability. In this paper we propose a Framework for Data Ecosystem Health Assessment (FASED) composed of indicators, which assess the current state and the ability to stay healthy over time. Each indicator uses metrics that allow the evaluation of the main elements and allows its continuous application to evaluate the health of the ecosystem under analysis. Using the proposed framework, it is expected to obtain indicators on the state of the indicators to assist in decision making related to the processes of publication and consumption of data, as well as to evaluate which ecosystems are more prosperous or need more investments.

Keywords: Framework. Health Assessment. Data Ecosystems.

LISTA DE FIGURAS

Figura 1 – Atores do Ecossistema de Dados do Twitter.	15
Figura 2 – Etapas da Metodologia de Pesquisa.	19
Figura 3 – Ciclo de Vida do <i>Design Science Research</i>	20
Figura 4 – Relação entre as classificações de Iansiti e Levien (2002) e Hartigh, Tol e Visscher (2006)	37
Figura 5 – Visão geral do framework OSEHO.	38
Figura 6 – Framework para gerenciar ECOs sustentáveis.	39
Figura 7 – Modelo de qualidade QuESo.	40
Figura 8 – Arquitetura HEAL ME.	41
Figura 9 – Modelo Twin Peaks.	45
Figura 10 – Estrutura do framework	46
Figura 11 – FASED - Hierarquia e Elementos	66
Figura 12 – Resultados da 1ª Avaliação - Métricas da Característica Consumo	77
Figura 13 – Resultados da 1ª Avaliação - Características do Indicador Produtividade	78
Figura 14 – Resultados da 1ª Avaliação - Características do Indicador Robustez	79
Figura 15 – Resultados da 1ª Avaliação - Características do Indicador Criação de Nicho	79
Figura 16 – Resultados da 1ª Avaliação - Características do Indicador Sustentabi- lidade	80
Figura 17 – Resultados da 1ª Avaliação - Visão Geral dos Indicadores	81
Figura 18 – Resultados da 1ª Avaliação - Saúde Geral do EDGov	81
Figura 19 – Resultados da 2ª Avaliação - Métricas da Característica Consumo	85
Figura 20 – Resultados da 2ª Avaliação - Características do Indicador Produtividade	86
Figura 21 – Resultados da 2ª Avaliação - Características do Indicador Robustez	86
Figura 22 – Resultados da 2ª Avaliação - Características do Indicador Criação de Nicho	87
Figura 23 – Resultados da 2ª Avaliação - Características do Indicador Sustentabi- lidade	87
Figura 24 – Resultados da 2ª Avaliação - Visão Geral dos Indicadores	88
Figura 25 – Resultados da 2ª Avaliação - Saúde Geral do EDGov	88
Figura 26 – Avaliação do Estado da Saúde - Métricas da Característica Gestão de Riscos	90
Figura 27 – Avaliação do Estado da Saúde - Métricas da Característica Consistência Financeira	91
Figura 28 – Avaliação do Estado da Saúde - Indicadores	92
Figura 29 – Avaliações da Saúde em Períodos de Tempo Distintos	92

Figura 30 – Resultado da Avaliação dos Indicadores - Pergunta 1	97
Figura 31 – Resultado da Avaliação dos Indicadores - Pergunta 2	98
Figura 32 – Resultado da Avaliação das Características de Produtividade - Pergunta 1	100
Figura 33 – Resultado da Avaliação das Características de Produtividade - Pergunta 2	100
Figura 34 – Resultado da Avaliação das Características de Robustez - Pergunta 1 .	101
Figura 35 – Resultado da Avaliação das Características de Robustez - Pergunta 2 .	102
Figura 36 – Resultado da Avaliação das Características de Criação de Nicho - Pergunta 1	103
Figura 37 – Resultado da Avaliação das Características de Criação de Nicho - Pergunta 2	103
Figura 38 – Resultado da Avaliação das Características de Sustentabilidade - Pergunta 1	105
Figura 39 – Resultado da Avaliação das Características de Sustentabilidade - Pergunta 2	105
Figura 40 – Resultado da Avaliação das Métricas de Produtividade - Pergunta 1 . .	106
Figura 41 – Resultado da Avaliação das Métricas de Produtividade - Pergunta 2 . .	107
Figura 42 – Resultado da Avaliação das Métricas de Robustez - Pergunta 1	107
Figura 43 – Resultado da Avaliação das Métricas de Robustez - Pergunta 2	108
Figura 44 – Resultado da Avaliação das Métricas de Criação de Nicho - Pergunta 1	109
Figura 45 – Resultado da Avaliação das Métricas de Criação de Nicho - Pergunta 2	110
Figura 46 – Resultado da Avaliação das Métricas de Sustentabilidade - Pergunta 1 .	111
Figura 47 – Resultado da Avaliação das Métricas de Sustentabilidade - Pergunta 2 .	112

LISTA DE TABELAS

Tabela 1 – Lista dos Estudos Selecionados	23
Tabela 2 – Papéis dos Atores em Ecossistemas de Dados	27
Tabela 3 – Estudos que Descrevem a Estrutura Organizacional de Ecossistemas de Dados	29
Tabela 4 – Métricas do Indicador de Produtividade para Avaliar a Saúde de EDs .	55
Tabela 5 – Métricas do Indicador de Robustez para Avaliar a Saúde de EDs	57
Tabela 6 – Métricas do Indicador de Criação de Nicho para Avaliar a Saúde de EDs	60
Tabela 7 – Métricas do Indicador de Sustentabilidade para Avaliar a Saúde de EDs	62
Tabela 8 – Aplicação do FASED no EDGov - 1 ^a Avaliação	74
Tabela 9 – Aplicação do FASED no EDGov - 2 ^a Avaliação	81
Tabela 10 – Informações sobre os Participantes do Grupo Focal	95
Tabela 11 – Resultado das Perguntas Gerais sobre o FASED	113

SUMÁRIO

1	INTRODUÇÃO	14
1.1	MOTIVAÇÃO	14
1.2	CARACTERIZAÇÃO DO PROBLEMA E QUESTÃO DE PESQUISA	16
1.3	JUSTIFICATIVA	17
1.4	OBJETIVOS	17
1.5	METODOLOGIA DE PESQUISA	18
1.6	ESTRUTURA DA DISSERTAÇÃO	21
2	ECOSSISTEMAS DE DADOS	22
2.1	QUANTO À CARACTERIZAÇÃO	23
2.2	QUANTO AOS ELEMENTOS	24
2.3	QUANTO AOS ATORES E PAPÉIS	25
2.4	QUANTO À ESTRUTURA ORGANIZACIONAL	28
2.5	QUANTO À INFRAESTRUTURA	29
2.6	QUANTO À CRIAÇÃO DE VALOR	29
2.7	QUANTO AO GERENCIAMENTO DO ECOSISTEMA	30
2.8	QUANTO À PRIVACIDADE DOS DADOS	32
2.9	QUANTO À BASE TEÓRICA	32
2.10	QUANTO AOS BENEFÍCIOS E DESAFIOS	33
3	TRABALHOS RELACIONADOS	35
3.1	DEN HARTIGH ET AL. 2006	35
3.2	JANSEN 2014	37
3.3	DHUNGANA ET AL. 2010	38
3.4	FRANCO-BEDOYA ET AL. 2014	39
3.5	CARVALHO ET AL. 2017	40
3.6	SÍNTESE DO CAPÍTULO	41
4	FASED: FRAMEWORK PARA AVALIAÇÃO DA SAÚDE DE ECOSISTEMAS DE DADOS	43
4.1	METODOLOGIA PARA CONSTRUÇÃO DO FRAMEWORK	44
4.2	FASED: FRAMEWORK PARA AVALIAÇÃO DA SAÚDE DE ECOSISTEMAS DE DADOS	46
4.2.1	Indicadores	47
4.2.2	Características	49
4.2.3	Atributos	51

4.2.4	Métricas	52
4.2.4.1	Métricas de Produtividade	54
4.2.4.2	Métricas de Robustez	56
4.2.4.3	Métricas de Criação de Nicho	60
4.2.4.4	Métricas de Sustentabilidade	61
4.3	SÍNTESE DO CAPÍTULO	67
5	GUIA DE INSTANCIÇÃO DO FASED	68
5.1	COMO REALIZAR UMA AVALIAÇÃO	68
5.1.1	Processo de Avaliação	68
5.2	CENÁRIO MOTIVACIONAL E INSTANCIÇÃO DO FASED	70
5.2.1	Aplicação do FASED no EDGOV - 1ª Avaliação	73
5.2.2	Resultados da 1ª Avaliação da Saúde do EDGOV	76
5.2.3	Aplicação do FASED no EDGOV - 2ª Avaliação	81
5.2.4	Resultados da 2ª Avaliação da Saúde do EDGOV	84
5.2.5	Estado da Saúde em Períodos Distintos	89
5.3	SÍNTESE DO CAPÍTULO	93
6	AVALIAÇÃO DO FASED	94
6.1	GRUPO FOCAL	94
6.2	ORGANIZAÇÃO DO GRUPO FOCAL	94
6.2.1	Planejamento	94
6.2.2	Condução do Grupo Focal	96
6.2.3	Análise dos Dados	97
6.2.3.1	Avaliação dos Indicadores	97
6.2.3.2	Avaliação das Características do Indicador de Produtividade	99
6.2.3.3	Avaliação das Características do Indicador de Robustez	101
6.2.3.4	Avaliação das Características do Indicador de Criação de Nicho	102
6.2.3.5	Avaliação das Características do Indicador de Sustentabilidade	104
6.2.3.6	Avaliação das Métricas do Indicador de Produtividade	105
6.2.3.7	Avaliação das Métricas do Indicador de Robustez	107
6.2.3.8	Avaliação das Métricas do Indicador de Criação de Nicho	109
6.2.3.9	Avaliação das Métricas do Indicador de Sustentabilidade	110
6.2.3.10	Perguntas Gerais	113
7	CONCLUSÃO	115
7.1	CONSIDERAÇÕES FINAIS	115
7.2	LIMITAÇÕES	116
7.3	TRABALHOS FUTUROS	117

REFERÊNCIAS 118

0em

APÊNDICE A – FORMULÁRIO DE AVALIAÇÃO DO FASED . . . 124

1 INTRODUÇÃO

Este capítulo tem o objetivo de introduzir a pesquisa sobre saúde de Ecossistemas de Dados. Ele é composto pela motivação, caracterização do problema, a questão que iremos responder ao longo da dissertação, os objetivos gerais e específicos, justificativa e a metodologia de pesquisa.

1.1 MOTIVAÇÃO

Nos últimos anos, estudos voltados para ED vêm crescendo por causa da relevância que os dados apresentam nas tomadas de decisão em diferentes contextos, desde ambientes governamentais até ambientes corporativos. O rápido crescimento das redes de comunicação, Internet das Coisas, do inglês *Internet of Things* (IoT), e tecnologias *Web* relacionadas, revelam novas possibilidades de captura, armazenamento, publicação e análise dos dados (CHEN; MAO; LIU, 2014). Governos, instituições de pesquisa e indivíduos estão produzindo e disponibilizando grandes quantidades de dados em diversos tipos de plataformas (*e.g.* na *Web*, aplicações baseadas em sensores e mídias sociais) (CHEN; MAO; LIU, 2014). Cada vez mais os indivíduos estão reconhecendo a importância dos dados e estão criando plataformas e serviços para publicar, negociar ou vender dados (BARBOSA et al., 2014) (OLIVEIRA et al., 2016).

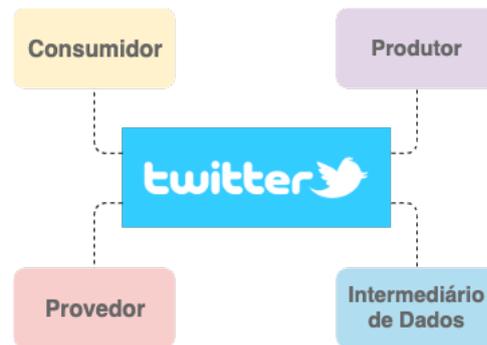
De acordo com Pollock (2011), na maioria desses casos, o modelo básico de disponibilização e uso dos dados é uma via de mão única. Não há *feedback* entre os produtores e os consumidores de dados, *i.e.* os consumidores dos dados não compartilham dados e informações úteis de volta para os produtores de dados. Em um cenário ideal, os consumidores de dados compartilhariam de volta seus dados refinados e integrados, por exemplo. Os consumidores de dados também deveriam ser capazes de compartilhar com os produtores os erros encontrados, ou submeter correções feitas por eles. De fato, todos os consumidores e produtores deveriam ser capazes de colaborar como meio para gerar valor e alavancar o crescimento do ecossistema.

Neste contexto, um ED pode ser visto como “*uma rede de atores composta por atores autônomos que diretamente ou indiretamente consomem, produzem ou fornecem dados e outros recursos relacionados a dados (e.g. software, serviços e infraestrutura). Cada ator desempenha um ou mais papéis e está conectado a outros atores por meio de relacionamentos, de forma que a colaboração e competição entre os atores promove a auto-regulação do Ecossistema de Dados*” (OLIVEIRA; LÓSCIO, 2018).

No contexto de dados de mídias sociais, o ED do Twitter possui atores e papéis que se relacionam de forma complexa, pois esse ecossistema vivencia um rápido crescimento no volume de dados, diferentes interesses e mudanças nas políticas de uso e compartilha-

mento. O ED do Twitter, Figura 1, é composto por usuários finais privados e institucionais, como cidadãos, governo e pesquisadores, os quais consomem os dados de forma direta e/ou indireta, por meio de soluções que utilizam esses dados. Também se destacam nesse ED os atores intermediários que coletam os dados com o intuito de gerar lucro por meio de novos recursos de compartilhamento de informações. De posse dessas informações, surge uma outra categoria de atores, cientistas de dados, os quais realizam a análise dos dados coletados e geram *insights* e previsões para diferentes áreas, como mercado financeiro, comportamentos sociais, saúde, educação, entre outros. E, por fim, o impulsionador do ecossistema e provedor dos dados, o próprio Twitter. Essa plataforma agrega atores com diferentes objetivos e facilita tanto o consumo quanto a produção de novos recursos, por meio de APIs, aplicações Web e mobile, e políticas de regulação dos dados (PUSCHMANN; BURGESS, 2013).

Figura 1 – Atores do Ecossistema de Dados do Twitter.



Fonte: A autora (2019)

Enquanto os EDs estão ganhando importância pelo seu potencial, alguns ainda não conseguem se manter produtivos, trazer novos investimentos, entregar valor e, por isso, não sobrevivem por longos períodos de tempo. Consequentemente, o esforço investido pelos seus atores acaba não sendo bem aproveitado ou até esquecido. A falta de comunicação e cooperação entre os produtores e consumidores de dados, é também um dos principais obstáculos para manter um Ecossistema de Dados (GAMA; LÓSCIO, 2014). Além disso, planejar, desenvolver e manter soluções para EDs não é trivial. Isso é causado pela dimensão e complexidade dos EDs, e a identificação de aspectos que impactam negativamente nesses ecossistemas se torna uma tarefa desafiadora. Sendo assim, como forma de identificar deficiências no funcionamento e atender às expectativas dos participantes é necessário ter indicadores que reflitam a saúde dos EDs. Saúde é um termo da biologia e se refere aos *status* do sistema ou de espécies específicas. Como nos Ecossistemas Naturais, a saúde dos EDs fornece informações sobre o estado atual dos seus componentes e sobre a sua capacidade de crescer e sobreviver ao longo do tempo.

Por se tratar de um tema recente, não existem trabalhos que proponham a operacionalização de métricas para avaliar a saúde dos EDs (OLIVEIRA; LIMA; LÓSCIO, 2019). Até

agora, alternativas para medir a saúde de EDs ainda são ingênuas, com foco em métricas relativamente simplistas, como número de conjuntos de dados publicados, número e porcentagem de conjuntos de dados baixados, número de conjuntos de dados programados para lançamento, número de APIs e análise básica do site (por exemplo, número de visualizações de páginas, downloads, etc.) (DAWES; VIDIASOVA; PARKHIMOVICH, 2016).

De maneira geral, a avaliação da saúde de ecossistemas é citada na literatura como uma forma de produzir indicadores sobre as atividades do ecossistema e os status dos elementos que o compõem (OLIVEIRA; LIMA; LÓSCIO, 2019). Os indicadores são obtidos por meio de métricas que avaliam o estado atual da saúde do ecossistema e a capacidade de permanecer saudável ao longo do tempo. Na literatura de saúde de ecossistemas, Costanza (1992) define um ecossistema saudável como a sua habilidade em manter a estrutura (organização) e função (vigor) ao longo do tempo mesmo ao enfrentar forças externas que gerem estresse (resiliência). Esse estudo e outros nas áreas de saúde de ecossistemas, como (IANSITI; LEVIEN, 2002) e (DHUNGANA et al., 2010), serviram de base para os trabalhos de avaliação da saúde de ecossistemas de negócios e *software*, como (MANIKAS; HANSEN, 2013b) (JANSEN, 2014) (CARVALHO et al., 2017a).

1.2 CARACTERIZAÇÃO DO PROBLEMA E QUESTÃO DE PESQUISA

São características inerentes a um ED a diversidade de dados e soluções baseadas em dados que trazem desafios na produção, consumo e manutenção desses recursos. Além disso, a heterogeneidade dos atores e papéis, assim como a complexidade dos relacionamentos, geram obstáculos aos envolvidos no gerenciamento do ecossistema. Por isso, é essencial a adoção de estratégias que auxiliem no funcionamento e propiciem longevidade ao ecossistema. Com base nesses fatores e levando em consideração os atores (*e.g.* empresas, desenvolvedores) que investem dinheiro e se associam às atividades do ecossistema para estimular o seu crescimento, é importante garantir que este esteja saudável. Desta forma, avaliar a saúde e gerar insumos sobre o estado da saúde do ED se revelam essenciais para a sua sobrevivência, entrega de valor e manutenção da sua produtividade. De acordo com Harrison, Pardo e Cook (2012), parte do que define a efetividade e sucesso de um ecossistema é baseado na definição e utilização de métricas que meçam a sua saúde.

Por outro lado, soluções para avaliar a saúde de EDs não são consolidadas na literatura. Diversos trabalhos avaliam a saúde de ecossistemas em outras áreas (*software*, negócios), mas eles não se aplicam em sua totalidade à realidade heterogênea e dinâmica dos EDs, e não especificam com detalhes como aplicar as métricas e interpretar os resultados. Como consequência, existe uma escassa base conceitual na área de saúde de EDs e poucas soluções que podem ser usadas como referência. Então, o problema apresentado surge como uma oportunidade para aumentar a contribuição acadêmica nas áreas de EDs e avaliação da saúde de ecossistemas. Portanto, este trabalho tem o intuito de contribuir

com um melhor entendimento sobre avaliação da saúde de EDs e como essa avaliação pode ser aplicada na prática.

Considerando a motivação apresentada e com base no problema descrito, nossa questão de pesquisa é: **Como a saúde de Ecossistemas de Dados pode ser avaliada?**

1.3 JUSTIFICATIVA

Pela dimensão e diversidade inerente aos EDs e por ser uma área que começou a ser explorada em 2011 (ZELETI; OJO, 2016) (DING et al., 2011), a literatura atual de EDs apresenta uma carência de trabalhos que abordem a avaliação da saúde desses ecossistemas. Em acréscimo, a especificação de como aplicar e interpretar as métricas de avaliação de saúde são inexistentes em EDs e escassas em outros domínios. Com base nesses motivos, este trabalho propõe um *framework* para avaliação da saúde de EDs, denominado FASED, consistindo de um conjunto de indicadores, características e métricas que permitem avaliar o ED baseado nos seus principais constructos (*e.g.* atores, relações, recursos e papéis). Com a aplicação das métricas e a análise dos resultados obtidos, é possível gerar indícios sobre o estado de cada indicador e assim ter um panorama da saúde do ED. A proposta do FASED foi idealizada e desenvolvida para contribuir com a área de Avaliação da Saúde de Ecossistemas, Ecossistemas de Dados e como forma de tratar os desafios encontrados nesta área.

1.4 OBJETIVOS

O objetivo principal deste trabalho é propor o *framework* FASED, um *framework* capaz de avaliar a saúde de Ecossistemas de Dados, utilizando indicadores e métricas de caráter genérico com o intuito de ser adaptável e aplicável em qualquer domínio de EDs.

Como objetivos específicos destacam-se:

- Realização de um mapeamento sistemático na área de Ecossistemas de Dados para definir o estado da arte;
- Revisão *ad-hoc* de trabalhos que avaliam a saúde de ecossistemas em outras áreas e dos trabalhos que eles referenciam;
- Construção do *framework* seguindo a abordagem *top-down*, partindo dos indicadores até as métricas;
- Especificação das métricas por meio de descrição, fórmula, interpretação dos resultados, unidade de medida, atributo do ecossistema relacionado e a referência bibliográfica;
- Avaliação do *framework* proposto.

1.5 METODOLOGIA DE PESQUISA

Normalmente, vários métodos podem ser aplicados aos problemas de pesquisa, e, em muitos casos, é necessário uma combinação de métodos para um entendimento completo do problema. Para Easterbrook et al. (2008), um método de pesquisa é um conjunto de princípios organizacionais em torno dos quais dados empíricos são coletados e analisados. Similarmente, Marconi e Lakatos (2010) afirmam que o método de pesquisa deve estar diretamente relacionado ao problema a ser estudado. Ou seja, uma pesquisa deve ser rigorosamente analisada mesmo antes de sua execução real.

Como diferentes metodologias de pesquisa servem propósitos diferentes, um dos primeiros passos é escolher um paradigma filosófico de pesquisa que seja apropriado (EASTERBROOK et al., 2008). O paradigma filosófico se preocupa com a fonte, natureza e desenvolvimento do conhecimento (CRESWELL, 2010). No geral, definir a filosofia de pesquisa envolve estar ciente e formular as expectativas e hipóteses da pesquisa. Nesse sentido, a proposta desta dissertação é baseada em um paradigma filosófico pragmático.

De acordo com Easterbrook et al. (2008), o pragmatismo valoriza o conhecimento prático sobre o conhecimento abstrato e usa todos os métodos apropriados para obtê-lo. Além disso, Easterbrook et al. (2008) e Creswell (2010) descrevem que essa postura filosófica é caracterizada pela aceitação de diferentes conceitos para apoiar a pesquisa. Desse modo, os pesquisadores que optam por esse paradigma filosófico acreditam que o problema é mais importante e usam todos os meios disponíveis para entendê-lo. Consequentemente, o paradigma pragmático não visa achar a verdade ou realidade, mas ajudar na resolução de problemas reais, buscando a solução para a questão em estudo, incluindo uma combinação de diferentes estratégias de pesquisa (CRESWELL, 2010).

Somado a isso, esta pesquisa também é baseada no paradigma de *Design Science Research*, o qual visa construir artefatos relevantes em termos de valor e de utilidade tanto em nível prático quanto teórico (HEVNER; CHATTERJEE, 2010). É fundamentalmente uma metodologia de resolução de problema que visa criar e avaliar artefatos de TI que resolvam um problema organizacional importante. De acordo com Alan et al. (2004), o artefato gerado durante a execução do ciclo de Design Science é um constructo, um modelo, um método ou uma instanciação. O *Design Science Research* é alinhado com a postura filosófica pragmática.

Segundo Hevner (2007), o artefato resultante deve possuir duas características essenciais: relevância e inovação. É necessário que o artefato resolva um problema importante de forma inovadora, eficiente e eficaz (GEERTS, 2011). E que seja descrito de forma efetiva para que permita sua implementação e aplicação em um domínio apropriado (ALAN et al., 2004). Venable, Pries-Heje e Baskerville (2012) acrescenta uma terceira característica: rigor. O processo envolvido na construção e avaliação do artefato deve ser rigoroso (VENABLE; PRIES-HEJE; BASKERVILLE, 2012). Conforme afirma Hevner (2007), é o rigor da construção do artefato que distingue Design Science Research da prática de concepção de

produtos na TI.

Portanto, com o auxílio do paradigma filosófico pragmático e do *Design Science Research*, este trabalho visa investigar e desenvolver um *framework* para avaliação da saúde de Ecossistemas de Dados. Devido à ausência de trabalhos nessa área e Ecossistemas de Dados ser um tema emergente, sob uma perspectiva objetiva, esta pesquisa é exploratória e descritiva. Exploratória porque visa fornecer um melhor entendimento do problema proposto e descritiva porque visa descrever um indivíduo, grupo ou fenômeno específico (ROBERT, 1994).

Em resumo, essa pesquisa foi realizada em quatro etapas, como mostra a Figura 2. Inicialmente, foi realizado um mapeamento sistemático da literatura com o intuito de obter uma visão geral da pesquisa na área de Ecossistemas de Dados. Segundo Keele et al. (2007), um mapeamento sistemático é uma metodologia orientada por protocolos que tem o objetivo de revisar e sintetizar trabalhos de uma área de pesquisa. Essa mapeamento teve como objetivo oferecer um snapshot das pesquisas na área de ED (i) identificando as diferentes definições de ED, (ii) analisando a evolução da pesquisa na área, (iii) analisando como os EDs são estruturados e organizados e (iv) identificando os benefícios e limitações dos EDs. Esta etapa foi de suma importância, pois formou um embasamento teórico inicial para a continuidade da pesquisa e, a partir dos trabalhos analisados, foi possível identificar que, até então, não há trabalhos que propõem um *framework* para avaliação da saúde de Ecossistemas de Dados. Os resultados desse mapeamento foram publicados em (OLIVEIRA; LIMA; LÓSCIO, 2019).

Figura 2 – Etapas da Metodologia de Pesquisa.

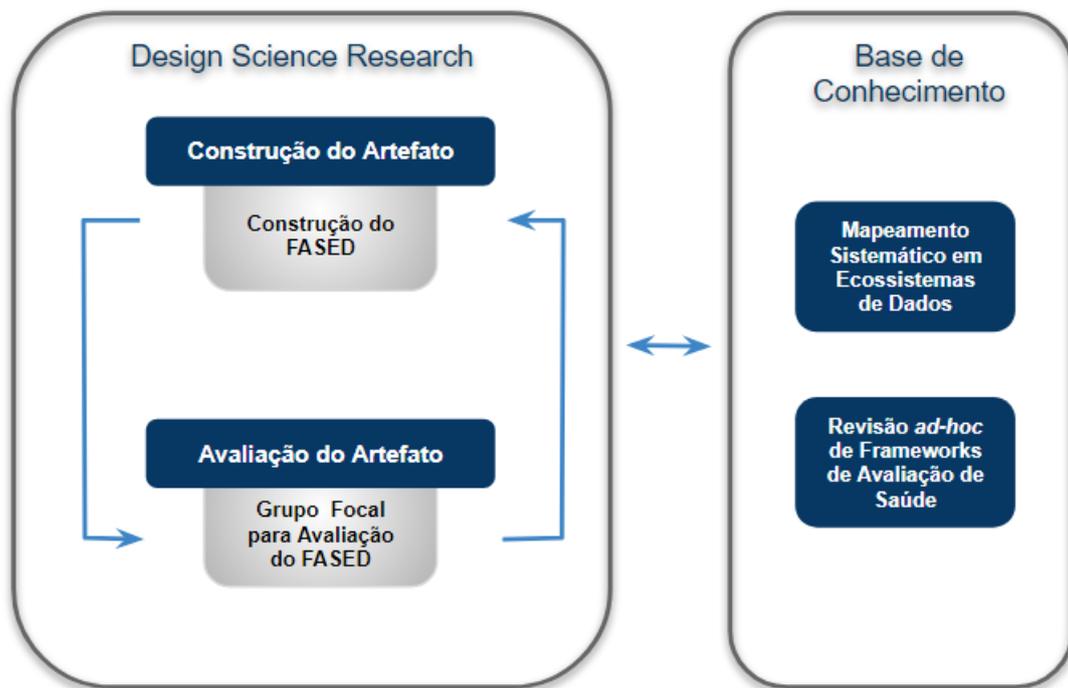


Fonte: A autora (2019)

A segunda etapa foi caracterizada por uma revisão *ad-hoc* da literatura para encontrar trabalhos relacionados ao nosso que pudessem nos fornecer um embasamento teórico. Uma revisão *ad-hoc* é uma busca informal na literatura em um dado assunto ou tópico de uma área específica. Em particular, uma revisão *ad-hoc* da literatura emprega métodos informais ou subjetivos para coletar e interpretar os estudos. Portanto, a revisão nos ajudou a identificar como os principais *frameworks* para avaliação de saúde são estruturados e aplicados. Essa etapa foi importante porque formou a base teórica inicial para o decorrer desta pesquisa. Essa revisão resultou na Seção 3.

Após a realização das pesquisas, a terceira etapa teve como objetivo construir e melhorar continuamente o *framework* proposto nesta dissertação. Essa etapa passou por vários ciclos de verificação e refinamento até ter como resultado uma versão inicial do *Framework* para Avaliação da Saúde de Ecossistemas de Dados. Por último, uma avaliação utilizando o método de grupo focal foi realizada. Um grupo focal pode ser definido como um grupo de indivíduos reunidos para avaliar produtos, sistemas, conceitos ou para evidenciar problemas (KONTIO; LEHTOLA; BRAGGE, 2004). Com base nos resultados obtidos nesta avaliação, o modelo foi refinado a fim de atender as melhorias propostas pelos participantes e sua versão final é apresentada no Capítulo 4.

Figura 3 – Ciclo de Vida do *Design Science Research*.



Fonte: A autora (2019)

Essas etapas também estão alinhadas com o ciclo de vida do *Design Science Research*. Figura 3 retrata o ciclo de vida do *Design Science Research* encontrado em (HEVNER, 2007) e sobrepõe os passos apresentados anteriormente. A caixa de *Design Science Rese-*

arch é o ciclo de Design que itera entre as atividades de construir e avaliar os artefatos. Já a caixa de Base de Conhecimento embasa a construção do artefato através de fundamentos científicos que informa o projeto de pesquisa.

1.6 ESTRUTURA DA DISSERTAÇÃO

Os próximos capítulos estão organizados da seguinte forma. No Capítulo 2 apresentamos a fundamentação teórica deste trabalho e no Capítulo 3 relatamos os trabalhos relacionados. Já no Capítulo 4 apresentamos o *framework* para avaliação da saúde de Ecossistemas de Dados, denominado FASED. No Capítulo 5 apresentamos um guia de instanciamento do FASED e aplicamos o *framework* duas vezes para avaliar a saúde de um Ecossistema de Dados Abertos Governamentais fictício, detalhamos os achados e mapeamos o estado da saúde após o seu uso. No Capítulo 6 apresentamos a técnica de grupo focal utilizada para avaliação do *framework*. Por fim, no Capítulo 7 é feita uma discussão sobre o trabalho realizado, suas limitações e sugestões para trabalhos futuros.

2 ECOSSISTEMAS DE DADOS

A maneira como indivíduos e organizações estão produzindo, compartilhando e consumindo dados mudou e está mudando com o advento de novas tecnologias. Como consequência, os dados se tornaram bens valiosos e estão fazendo parte das transações comerciais. Hoje, existem Ecosistemas de Dados nos quais comunidades de atores interagem umas com as outras para trocar, produzir e consumir dados, o que reflete na alta popularidade da área de Ecosistemas de Dados nos últimos anos.

Diversos fatores têm impulsionado o surgimento de Ecosistemas de Dados, incluindo o surgimento de tecnologias digitais e iniciativas políticas/institucionais. Por exemplo, a maioria dos ecossistemas de dados tem sido impulsionada principalmente pelos programas *Open Data Movement* e *Open Government Data* (OGD), que recomenda o livre uso, reutilização e redistribuição de dados por qualquer pessoa (*OPEN GOVERNMENT WORKING GROUP*, 2007). Vários governos já lançaram portais de dados abertos para estimular e promover a produção e o consumo dos dados (CHUN et al., 2010). Melhorias na tecnologia (por exemplo, dispositivos e internet móvel) e novas tendências (por exemplo, mídias sociais e aplicativos móveis) também têm impulsionado organizações públicas e privadas a publicar dados, assim como a integrar serviços com dados externos.

Neste capítulo, apresentamos uma visão geral da literatura atual sobre Ecosistemas de Dados, mostrando um subconjunto dos resultados encontrados no mapeamento sistemático publicado em (OLIVEIRA; LIMA; LÓSCIO, 2019) e os trabalhos analisados na Tabela 1. Em particular, este capítulo pretende apresentar um panorama das pesquisas na área de Ecosistema de Dados por meio da (I) definição e análise da sua estrutura e organização, (II) identificação dos principais atores e papéis, bem como sua estrutura organizacional, (III) mostrar como os Ecosistemas de Dados são gerenciados e como se dá a geração de valor nesse tipo de ecossistema e (IV) a apresentação dos fundamentos teóricos.

Tabela 1 – Lista dos Estudos Selecionados

ID do Estudo	Referência	ID do Estudo	Referência
S01	(SMITH; OFE; SANDBERG, 2016)	S16	(ZUIDERWIJK; JANSSEN; DAVIS, 2014)
S02	(MERCADO-LARA; GIL-GARCIA, 2014)	S17	(MAGALHAES; ROSEIRA; MANLEY, 2014)
S03	(ZELETI; OJO, 2016)	S18	(LEE, 2014)
S04	(MOISO; MINERVA, 2012)	S19	(BOURNE; LORSCH; GREEN, 2015)
S05	(SHIN, 2016)	S20	(ATTARD; ORLANDI; AUER, 2016)
S06	(HA; LEE; LEE, 2014)	S21	(LINDMAN; KINNARI; ROSSI, 2015)
S07	(ZUBCOFF et al., 2016)	S22	(IMMONEN; PALVIAINEN; OVASKA, 2014)
S08	(DING et al., 2011)	S23	(ZELETI; OJO, 2017)
S09	(ZUIDERWIJK et al., 2016)	S24	(ZUIDERWIJK et al., 2015)
S10	(DAWES; VIDIASOVA; PARKHIMOVICH, 2016)	S25	(HEIMSTÄDT; SAUNDERSON; HEATH, 2014)
S11	(DONKER; LOENEN, 2017)	S26	(HARRISON; PARDO; COOK, 2012)
S12	(KÖSTER; SUÁREZ, 2016)	S27	(DAVIES, 2011)
S13	(KOZNOV et al., 2016)	S28	(ZELETI; OJO, 2014)
S14	(SHIN; CHOI, 2015)	S29	(GAMA; LÓSCIO, 2014)
S15	(SCHALKWYK; WILLMERS; MCNAUGHTON, 2016)	-	-

Fonte: Oliveira, Lima e Lóscio (2019)

2.1 QUANTO À CARACTERIZAÇÃO

Ecossistemas de Dados podem ser definidos como uma comunidade de atores com iniciativas de compartilhamento de dados. Esses atores podem ser produtores de dados, usuários finais e intermediários, e interagem dinamicamente por meio da troca de recursos (por exemplo, dados, serviços, ferramentas e infraestrutura). Essas interações formam um tipo de relacionamento, algumas delas representam relações de curta ou longa duração.

De forma geral, um Ecossistema de Dados aborda a natureza de múltiplos aspectos de compartilhamento de dados, abrangendo vários contextos, como as sociais, legais, políticas, culturais, tecnológicas, operacionais e econômicas. Essa natureza abrangente enfatiza que todos os aspectos de um ecossistema estão interligados e são influenciados por valores (HARRISON; PARDO; COOK, 2012).

De acordo com Schalkwyk, Willmers e McNaughton (2016) existem pelo menos três condições contextuais que influenciam os elementos dos ecossistemas. O primeiro deles é o contexto regulatório, que inclui leis, políticas, normas e acordos. Este contexto orienta a especificação de como os elementos do ecossistema são estruturados e como eles se inter-relacionam. O segundo é o contexto institucional em que os atores operam. Cada contexto

institucional fornece valores, regras e normas que inevitavelmente impulsionam e restringem os comportamentos dos atores no ecossistema. O terceiro é o contexto tecnológico, que abrange os recursos de TI, os operadores de TI e outras tecnologias que conectam e interconectam os elementos dos EDs. O contexto ambiental, como elementos culturais, sociais e econômicos, também exerce uma influência sobre iniciativas de EDs (LEE, 2014).

2.2 QUANTO AOS ELEMENTOS

Em relação aos elementos do Ecossistema de Dados, os autores Mercado-Lara e Gil-Garcia (2014) apontam três domínios importantes de elementos: (I) políticas e práticas governamentais, (II) *innovators* (uma combinação de tecnologia, negócios e governo) e (III) usuários, sociedade civil e negócios. Shin e Choi (2015) também identificam como elementos-chave de um Ecossistema de Dados: (I) infraestrutura, (II) software e tecnologias, (III) serviços e aplicações, (IV) normas, (V) usuários, (VI) fatores sociais e culturais, (VII) governo e (VIII) indústria. Para criar e capturar valor, Zeleti e Ojo (2017) afirmam que os atores devem empregar conjuntos emergentes de capacidades, os quais são tipos de habilidades. Nos quais permitem que uma atividade seja realizada ou até mesmo que a produtividade de algumas atividades sejam melhoradas. Essas capacidades podem incluir uma variedade de habilidades e especialidades, como conhecimento geral sobre recursos do ED (por exemplo, sistemas e tecnologias), habilidades técnicas necessárias para gerenciar e processar dados (por exemplo, integração de dados e técnicas de mineração de dados) e experiência operacional para incorporar recursos relacionados a dados em processos institucionais e empresariais existentes (LEE, 2014).

Além das capacidades, os atores também necessitam de recursos adequados para fornecer e consumir dados. Exemplos comuns de recursos são conjuntos de dados, serviços, ferramentas, capital financeiro, bem como capital humano, equipamentos, materiais e tecnologia proprietária. Zuiderwijk et al. (2015) distinguem entre três categorias de recursos: recursos humanos, recursos de dados e recursos de tecnologia da informação. De forma geral, os dados não são apenas o recurso principal, mas também vários outros tipos de recursos são complementares e necessários para que os EDs funcionem adequadamente.

Mercado-Lara e Gil-Garcia (2014) enfatizam as relações de dependência nos Ecossistemas de Dados. De acordo com esses autores, todos os elementos do ecossistema estão interconectados de forma que quando um elemento é alterado, os efeitos podem ser sentidos em todo o sistema. De fato, os atores afetam e são afetados pela criação e entrega de recursos executados por outros atores (IMMONEN; PALVIAINEN; OVASKA, 2014). Além disso, os atores têm seus próprios interesses que podem gerar conflitos. Em particular, os consumidores de dados são fortemente influenciados pela decisão de um produtor de dados de não publicar ou atualizar um determinado dado ou pela alteração do formato em que os dados são publicados, podendo assim comprometer a qualidade dos dados ou alterar como ele pode ser usado (ZELETI; OJO, 2016).

2.3 QUANTO AOS ATORES E PAPÉIS

Em Ecossistemas de Dados, os atores são entidades autônomas que podem ser empresas, instituições ou indivíduos, que atuam no ecossistema desempenhando um ou vários papéis. Os atores são motivados por um conjunto de interesses (ZUIDERWIJK; JANSSEN; DAVIS, 2014) e geralmente têm um compromisso com o ecossistema, podendo ter um incentivo para serem ativos no ecossistema.

Um papel é uma função desempenhada por um ator no ecossistema. Está relacionado a uma posição e a um conjunto de deveres. Diversas funções podem ser identificadas nos EDs, como apresentado na Tabela 2. Os papéis mais citados são de consumidor e produtor de dados. No entanto, existem vários papéis adicionais, cada um com funções diferentes. Além disso, é possível encontrar dois papéis diferentes compartilhando a mesma função. Por exemplo, tanto os intermediários de dados quanto os produtores de dados fornecem dados. Em relação a terminologia de Ecossistema de Dados, há pouco consenso sobre os papéis dos atores e seus respectivos deveres e atividades. Segundo Lundell et al. (2009), definir os papéis em cada ecossistema é essencial para pesquisadores e analistas de negócios, a fim de compreender e gerenciar o ecossistema e poder estimar seu sucesso.

A função mais frequentemente identificada é o usuário de dados, responsável pelo consumo direto ou indireto de dados. Quase todos os estudos apresentam o conceito do papel de usuário de dados. Os usuários de dados não têm necessariamente a capacidade de consumir dados diretamente dos produtores de dados. Eles geralmente dependem de serviços fornecidos por usuários que criam valor a partir do reuso dos dados, intermediários de dados ou provedores de serviços. Além disso, os usuários de dados geralmente representam os usuários finais de um ecossistema.

O segundo papel mais destacado nos EDs é o produtor de dados, responsável pelo fornecimento de dados. É importante notar que os produtores de dados não são apenas responsáveis pela captura e geração de dados, mas também pela compilação, agregação e empacotamento de dados.

Outro papel identificado é o ator intermediário, responsável por usar ou transformar os dados em novos recursos, e também investir no ecossistema. Esses atores podem ser indivíduos ou organizações privadas ou públicas. Dentre as atividades desenvolvidas estão: desenvolvimento de aplicações, serviços, visualização de dados e fomento dos recursos do ecossistema.

O ator que exerce o papel de *keystone* é responsável por impulsionar o ecossistema, bem como fornecer estabilidade em situações de instabilidade (IANSITI; LEVIEN, 2004a). Esse papel é muito comum em ECOS. Em (DAWES; VIDIASOVA; PARKHIMOVICH, 2016), os autores afirmam que os *keystones* são responsáveis por fornecer a maioria dos dados, bem como por promover o ecossistema. Lee (2014) afirma que esse papel deve ser atribuído aos atores que lideram os programas governamentais de dados. Nesse cenário, os atores *keystones* devem promover um conjunto de diretivas, regras e práticas formais para condu-

zir um ED. Com uma visão ligeiramente diferente, Schalkwyk, Willmers e McNaughton (2016) afirmam que os *keystones* são facilitadores, não necessariamente condutores do ecossistema; eles são úteis, mas não são essenciais para o funcionamento contínuo de um ecossistema.

Outro papel citado na literatura é o de prestador de serviços. Eles são responsáveis por produzir e manter recursos de software, como ferramentas, aplicações, serviços, entre outros (GAMA; LÓSCIO, 2014). As partes de políticas, leis e regras representam o papel responsável pela criação de regras e políticas para incentivar e controlar a participação dos atores (ZELETI; OJO, 2016) (ZUIDERWIJK et al., 2016) (DONKER; LOENEN, 2017) (DAWES; VIDIASOVA; PARKHIMOVICH, 2016) (SCHALKWYK; WILLMERS; MCNAUGHTON, 2016). Normalmente, essa função é desempenhada por governos, fundadores de EDs ou instituições de padronização (por exemplo, *Open Knowledge Foundation* e W3C).

Outras funções importantes identificadas na literatura são o Provedor de Infraestrutura, Consultor de Dados, Patrocinador de Dados e Curador de Dados. Provedor de Infraestrutura é a função que suporta as atividades de outras funções. A infraestrutura inclui o fornecimento de recursos ou serviços de Tecnologias de Informação e Comunicação (TIC), como a capacidade de hospedagem ou armazenamento (HA; LEE; LEE, 2014) (LINDMAN; KINNARI; ROSSI, 2015) (IMMONEN; PALVIAINEN; OVASKA, 2014). Um Consultor de Dados auxilia outros papéis para analisar as possibilidades que os dados podem proporcionar e também identifica as necessidades do ator (IMMONEN; PALVIAINEN; OVASKA, 2014). Um Patrocinador de Dados é responsável por promover a iniciativa *Open Data* por meio de programas de financiamento público e investimentos privados (BOURNE; LORSCH; GREEN, 2015) (GAMA; LÓSCIO, 2014). Por fim, os Curadores de Dados são responsáveis pela qualidade e disponibilidade de dados (BOURNE; LORSCH; GREEN, 2015) (DING et al., 2011).

Tabela 2 – Papéis dos Atores em Ecossistemas de Dados

Papel	Estudos
Provedor de Dados	S01, S03, S05, S06, S07, S09, S11, S12, S13, S14, S15, S17, S19, S20, S21, S22, S23, S24, S25, S27, S29
Armazenador	S22
Desenvolvedor	S22
Agregador	S22, S29
Harmonizador	S22
Publicador	S22
Registrador	S22
Mantenedor	S19
Produtor de Dados	S02, S04, S08, S16, S19, S26,
Proprietário de Dados	S21
Responsável por Políticas, Leis e Regras	S03, S09, S11, S10, S15
Responsável por Políticas	S09, S11
Responsável por Padrão e Regulação	S03, S15
Ator <i>Keystone</i>	S10, S15, S19, S26
Fundador	S19
Provedor de Serviço	S05, S06, S14, S21, S22, S29
Provedor de <i>Data-as-a-Service</i>	S22
Provedor de <i>Analytics-as-a-Service</i>	S22
Desenvolvedor de Serviço de Dados	S01, S29
<i>Re-user</i>	S01, S02, S08, S12, S18, S17, S21, S22, S23, S26, S29
Organização Orientada a Dados Abertos	S23
Desenvolvedor de Aplicação	S08, S22
Intérprete	S21, S22
Intermediário de Dados	S01, S02, S04, S09, S15, S16, S17, S21, S23, S25, S26
Corretor de Dados	S01, S04, S22
Facilitador	S17, S23
Integrador	S17
Transformador e Extrator de Dados	S21
Analisador de Dados	S21
Visualizador de Dados	S21
Usuário de Dados/Consumidor de Dados	S01, S02, S03, S04, S05, S06, S07, S08, S09, S10, S11, S12, S13, S14, S15, S16, S17, S19, S20, S21, S22, S23, S24, S25, S26, S27, S28, S29
Curador de Dados	S08, S19
Provedor de Infraestrutura	S06, S21, S22
Financiador de Dados	S19, S29
Consultor de Dados	S01, S22

Fonte: Oliveira, Lima e Lóscio (2019)

2.4 QUANTO À ESTRUTURA ORGANIZACIONAL

Em um ED, cada ator é conectado a outros atores por um conjunto de interesses ou modelos de negócios. Toda a rede de relacionamentos pode seguir uma estrutura organizacional, variando de uma abordagem difusa *ad hoc* a uma abordagem centrada no *keystone* (HANSSEN; DYBÅ, 2012). Uma estrutura organizacional do ecossistema leva em conta a maneira como os atores estão conectados e as propriedades de seus relacionamentos, como a dependência de relacionamentos (MANIKAS; HANSEN, 2013b).

Nos estudos selecionados foram encontrados 5 diferentes abordagens de estruturas organizacionais apresentadas na Tabela 3: centrada no *keystone*, baseada no intermediário de dados, centrada na plataforma, baseada no mercado e orientada a modelo de negócios.

Na estrutura centrada no *keystone*, os atores são organizados em torno dele, que é direta ou indiretamente responsável por fornecer grande parte dos dados. No entanto, o *keystone* não tem controle completo sobre os outros atores, eles podem sair (ou entrar) do ecossistema a qualquer momento. Somado a isso, o *keystone* é responsável pelo monitoramento, avaliação, tomada de decisões e ações (HEIMSTÄDT; SAUNDERSON; HEATH, 2014).

A estrutura baseada nos intermediários depende da presença de atores intermediários para gerar valor a partir dos dados. Um intermediário de dados é um papel que facilita o uso de dados para outros atores. Portanto, em uma estrutura baseada no intermediário de dados, produtores de dados e usuários de dados (isto é, os dois extremos em uma cadeia de fornecimento) são organizados em torno dos intermediários. Alguns estudos reconhecem o papel crítico que os intermediários podem desempenhar nos ecossistemas (SCHALKWYK; WILLMERS; MCNAUGHTON, 2016) (IMMONEN; PALVIAINEN; OVASKA, 2014) (LINDMAN; KINNARI; ROSSI, 2015).

Na estrutura centrada na plataforma, a plataforma define a estrutura organizacional de um ecossistema de dados. De acordo com o Dawes, Vidiasova e Parkhimovich (2016), a plataforma fornece infraestrutura e serviços para suportar o fornecimento e o consumo de dados. Dawes, Vidiasova e Parkhimovich (2016) e Ding et al. (2011) enfatizam que os custos de fornecimento de dados são reduzidos quando os dados são liberados através da plataforma. A plataforma também pode atenuar problemas de interoperabilidade e usabilidade. As ferramentas de catálogo de dados abertos (por exemplo, CKAN2) são comuns, mas exemplos limitados de plataformas são usados para criar Ecossistemas de Dados (GAMA; LÓSCIO, 2014).

Em estruturas baseadas no mercado, os mercados fornecem as infraestruturas, modelos de negócios, regras e serviços necessários para transações de dados e software entre os atores (SMITH; OFE; SANDBERG, 2016). Em geral, os mercados abrangem uma plataforma técnica com capacidade de vincular produtores de dados e usuários de dados. Eles também permitem a venda de dados, serviços e aplicativos.

Apesar de não definir como os atores devem ser organizados, alguns estudos apre-

sentam modelos de negócios, que descrevem a lógica de como um ator cria, entrega e captura valor. Em particular, valor refere-se a qualquer benefício que um agente obtém do ED, como satisfação, utilidade, solução de problemas ou receita. Modelos de negócios comuns são *business-to-business*, *business-to-consumer* e *consumer-to-consumer* (HA; LEE; LEE, 2014). Segundo Christensen et al. (2014), o modelo de negócios é importante para fundamentar o custo, a receita e/ou a sustentabilidade do ecossistema de dados.

Tabela 3 – Estudos que Descrevem a Estrutura Organizacional de Ecossistemas de Dados

Estrutura do ED	Estudos
Centrado no Keystone	S07, S10, S12, S13, S15, S16, S18, S25
Baseado no Ator Intermediário	S02, S04, S15, S21, S22
Centrado na Plataforma	S08, S12, S27,S29
Baseado no Mercado	S01, S04
Orientado a Modelo de Negócio	S06, S19, S20, S21, S22, S23, S24, S28

Fonte: Oliveira, Lima e Lóscio (2019)

2.5 QUANTO À INFRAESTRUTURA

Várias infraestruturas foram desenvolvidas nos últimos anos para criar plataformas de Ecossistema de Dados para explorar o potencial dos dados, como portais de dados abertos, infraestrutura de dados abertos europeus¹, infraestruturas de estatísticas (por exemplo, Barometer²), soluções de publicação de dados (por exemplo CKAN, Junar³ e Socrata⁴). Outra alternativa é usar os serviços como blocos de construção para construir plataformas para ecossistemas de dados. Além disso, essas infraestruturas também promovem as interações entre os atores do Ecossistema de Dados e devem também permitir que produtos e serviços inovadores sejam criados em grande escala, usando formatos e ferramentas comuns.

A falta de tecnologias padronizadas e diretrizes dificultam a facilidade de uso e reduzem a produtividade. Por exemplo, soluções heterogêneas implicam que os atores devem lidar com diferentes APIs (*Application Programming Interfaces*) e métodos de acesso a dados, a heterogeneidade estrutural, sintática e semântica dos dados e a variabilidade dos níveis de qualidade e a diversidade nos tipos de metadados.

2.6 QUANTO À CRIAÇÃO DE VALOR

Nos EDs os atores são obrigados a empregar um conjunto de capacidades e recursos para catalisar o valor. De acordo com o Magalhaes, Roseira e Manley (2014), muitas

¹ <https://www.eudat.eu/>

² <https://opendatabarometer.org/>

³ <http://www.junar.com/>

⁴ <http://www.socrata.com/>

vezes o ônus é dos consumidores para extrair valor dos recursos disponíveis. Isso cria um problema, já que o consumidor não possui as habilidades necessárias (ZUIDERWIJK et al., 2012). Devido à essas barreiras, o valor não é criado por um único ator, mas sim por uma cadeia de valor (ou seja, por uma rede de atores). Uma cadeia de valor é um conjunto de atividades independentes de valor agregado que é usado para explorar um conjunto de recursos. Além disso, uma cadeia de valor consiste em diferentes atores conduzindo uma ou mais atividades (por exemplo, provisão de dados, curadoria de dados, análise de dados), e cada atividade pode consistir em várias ações ou técnicas de criação de valor (por exemplo, coleta, visualização, criação de serviço). Em Ecossistemas de Dados, a cadeia de valor mínima consiste em produtores de dados, intermediários de dados e consumidores de dados (HEIMSTÄDT; SAUNDERSON; HEATH, 2014). Como as atividades que agregam valor possuem complexidade diferente, é possível que cada ação possa consistir em uma ou mais cadeias de valor (ATTARD; ORLANDI; AUER, 2016).

Os modelos de negócios também geram valor para os atores do ecossistema. Vários modelos de negócios aplicáveis a dados foram descritos na literatura (TEECE, 2010) (ZOTT; AMIT, 2010), como modelos de suporte, subscrição e serviços profissionais/consultoria, que poderiam ser aplicáveis tanto para produtores de dados quanto para desenvolvedores de aplicativos (IMMONEN; PALVIAINEN; OVASKA, 2014).

Descobrir como gerar receita (ou seja, obter lucro financeiro) ao fornecer dados ou recursos baseados em dados aos consumidores é um elemento-chave para criação de modelos de negócios em Ecossistemas de Dados. Serviços e aplicativos podem ser cotados comumente com base em modelos de recursos, custo ou *pay-per-use*, por exemplo (IMMONEN; PALVIAINEN; OVASKA, 2014). Os recursos de dados podem ser precificados usando o modelo de subscrição, no qual o consumidor paga um preço fixo por um determinado período de tempo. No entanto, em geral, os dados costumam ser complexos, e os consumidores têm muitas maneiras de obter certos tipos de dados sem pagar (TEECE, 2010).

2.7 QUANTO AO GERENCIAMENTO DO ECOSISTEMA

Embora os Ecossistemas de Dados tenham o potencial de gerar benefícios, muitas iniciativas de ecossistemas estão lutando para estabelecer uma gestão eficaz de seus recursos e atores (POLLOCK, 2011) (UBALDI, 2013). Por exemplo, existem preocupações fundamentais relacionadas ao desenvolvimento de Ecossistemas de Dados, apesar das altas expectativas e do investimento financeiro. De acordo com o Dawes, Vidiasova e Parkhimovich (2016), como consequência de inúmeras barreiras e limitações, o desempenho das iniciativas de Ecossistema de Dados tende a ser simplista. Eles focam em lançar e promover concursos de curta duração, como *hackathons*. A maioria das aplicações nesses cenários acaba sendo rapidamente abandonada. Em uma perspectiva de longo prazo, pouquíssimas aplicações resultantes de tais concursos são realmente escaláveis e sustentáveis (GAMA; LÓSCIO, 2014).

Outros elementos essenciais em um EDs de sucesso são a colaboração de atores, a integração de informações científicas, sociais e econômicas, a preservação de processos ecológicos e o gerenciamento (ZUIDERWIJK; JANSSEN; DAVIS, 2014). De fato, o gerenciamento de EDs é importante para facilitar ativamente o funcionamento efetivo e cumprir as metas dos ecossistemas (MERCADO-LARA; GIL-GARCIA, 2014). Além disso, se um Ecossistema de Dados não possui uma estrutura de gerenciamento, torna-se difícil impulsionar o ecossistema, e construir e aprender com experiências passadas (LEE, 2014).

De acordo com Mercado-Lara e Gil-Garcia (2014), o gerenciamento de um Ecossistema de Dados requer o esboço de alguns tópicos básicos, que se concentram em (I) identificar os atores mais ativos que atuam como componentes essenciais do ecossistema; (II) analisar a natureza das transações que ocorrem entre esses atores; (III) reconhecer quais recursos são necessários para cada ator e como eles realizam as transações; e (IV) estudar os indicadores que sinalizam o status da atividade do ecossistema (MERCADO-LARA; GIL-GARCIA, 2014). Portanto, essas considerações exigem uma abordagem mais sistêmica para programar o planejamento e projetar os Ecossistemas de Dados. No entanto, até o momento, as iniciativas de gerenciamento de ecossistemas de dados são simplistas. Governos de todo o mundo desenvolveram programas e políticas que visam promover o fornecimento e o uso de dados do setor público (MERCADO-LARA; GIL-GARCIA, 2014) (KOZNOV et al., 2016). Tais políticas muitas vezes se concentram em garantir a disponibilidade e a qualidade dos recursos de dados. Essas políticas de gerenciamento ajudaram a expandir os EDs e melhorar o fornecimento de dados. No entanto, essas políticas não incluem outros atores, como consumidores de dados e intermediários de dados, que realmente demandam o fornecimento. Por isso, é crucial incluir, desde o início, o ponto de vista de todos os atores do ecossistema. Uma gestão integrada e colaborativa deve assegurar que as metas incluídas em uma agenda de Ecossistema de Dados atendam às necessidades, direitos e interesses de todos os atores que fazem parte do ecossistema (KÖSTER; SUÁREZ, 2016).

Os métodos de avaliação são ferramentas úteis para melhorar a eficácia dos Ecossistemas de Dados. A literatura refere-se ao conceito de saúde de um ecossistema como meio de monitorar e avaliar o funcionamento do ecossistema, identificar e prever áreas de melhoria e avaliar mudanças no ecossistema (MANIKAS; HANSEN, 2013a). Até agora, alternativas para medir a saúde do ecossistema de dados ainda são ingênuas, como o foco em métricas relativamente simplistas, como número de conjuntos de dados publicados, número e porcentagem de conjuntos de dados baixados, número de conjuntos de dados programados para lançamento, número de APIs e análise básica do site. (por exemplo, número de visualizações de páginas, downloads, etc.) (DAWES; VIDIASOVA; PARKHIMOVICH, 2016). Embora informativos, até certo ponto, esses indicadores de saúde se concentram apenas na perspectiva do produtor de dados e, como consequência, não avaliam a utilidade dos recursos fornecidos.

2.8 QUANTO À PRIVACIDADE DOS DADOS

Apesar do funcionamento e impactos benéficos esperados dos EDs, governos, empresas privadas e indivíduos estão cientes dos riscos de privacidade, e estão começando a reivindicar maior proteção em seus dados (MOISO; MINERVA, 2012) (SHIN, 2016) (SHIN; CHOI, 2015). A cada dia, uma quantidade crescente de dados pessoais é coletada, analisada e usada por entidades privadas e públicas. Infelizmente, nem todos os detentores de dados pessoais oferecem garantia total aos seus usuários sobre o controle da privacidade dos dados. Conseqüentemente, na realização dos atos ilícitos e do uso indevido de dados, os governos e as organizações internacionais estão criando regulamentos, incluindo penas mais severas para pessoas ou grupos que publicam, sem justificativa, na maioria das vezes com o objetivo de prejudicar a reputação de outra pessoa, expondo-a ao ódio, desprezo ou ridicularização.

Como uma grande quantidade de dados é privada (por exemplo, dados pessoais), o gerenciamento da privacidade dos dados é importante (IMMONEN; PALVIAINEN; OVASKA, 2014). A privacidade de dados é um elemento que deve ser cuidadosamente considerado ao manipular quaisquer dados. Um conjunto de práticas deve ser implementada para garantir que as leis de proteção de dados sejam cumpridas. Além disso, há uma crescente preocupação para evitar que os dados pessoais não sejam vinculados a um indivíduo. A única exceção a isso é a informação sobre pessoas que estão legalmente no domínio público para fins de transparência. Em todos os casos de privacidade de dados é um direito fundamental que deve ser protegido e o publicador de dados deve estar em conformidade com as leis de proteção de dados.

Algumas medidas de mitigação podem ser empregadas para proteger a privacidade dos dados. Por exemplo, o fornecimento de diretrizes de privacidade claras para publicação de dados, contato com a autoridade nacional de estatísticas sobre melhores práticas de anonimato dos dados e com o escritório de proteção de dados para orientação sobre legislação cria uma campanha de conscientização para esclarecer diferenças entre diferentes tipos de publicação de dados. Antes de decidir quais dados publicar, os produtores de dados precisam ter uma visão geral dos dados que gerenciam. Essa pode ser uma tarefa desafiadora, pois os dados em grandes organizações geralmente são dispersos em vários sites, bancos de dados, armazenamento compartilhado e computadores pessoais.

2.9 QUANTO À BASE TEÓRICA

Um ED pode ser visto como outra instância de um Ecossistema de Negócios, um Ecossistema de Negócios Digital ou um Ecossistema de Software. Apesar de compartilharem características de rede e coevolução, os Ecossistemas de Dados diferem dos ecossistemas anteriores. Ao contrário de outros ecossistemas, os Ecossistemas de Dados não dependem de uma plataforma explícita comum na qual diferentes atores possam colaborar. A pla-

taforma comum é, na verdade, um grande conjunto de dados trocados pelos atores. Em particular, os dados não precisam necessariamente ser fornecidos por um único ator. A falta de uma plataforma comum cria uma rede de oferta e demanda mais difusa. Outra diferença está relacionada a como os produtos negociados entre os atores são percebidos. Em Ecossistemas de Negócios, as operações de negócios e os atores em si são os produtos (MANIKAS; HANSEN, 2013a). Em ecossistemas de software, os produtos são componentes ou serviços de software. Em Ecossistemas de Dados, o produto é os dados.

Na verdade, as fronteiras entre diferentes tipos de ecossistemas são difíceis de definir (ZUIDERWIJK; JANSSEN; DAVIS, 2014). Por exemplo, os Ecossistemas de Dados podem envolver os ecossistemas de software em relação à rede de atores envolvidos no desenvolvimento e fornecimento de software relacionado a dados. Assim, a compreensão desses ecossistemas relacionados nos permite usar alguns conceitos no mundo de Ecossistema de Dados. Ele pode nos fornecer soluções para o desafio de construir informações e conhecimento sobre um Ecossistema de Dados.

Em resumo, um conhecimento teórico heterogêneo tem sido utilizado pelos estudos de pesquisa do Ecossistema de Dados. Esta situação é uma consequência de dois fatores: a área está em sua fase inicial e diferentes comunidades de pesquisa e indústria têm investigado a área de forma independente.

2.10 QUANTO AOS BENEFÍCIOS E DESAFIOS

Falando sobre os benefícios dos EDS, os mais citados estão relacionados a possibilitar ou melhorar aspectos da vida política e social, como melhorias na qualidade de vida e confiança social, crescimento econômico, apoio a processos de formulação de políticas e melhoria dos serviços aos cidadãos (SMITH; OFE; SANDBERG, 2016) (DAWES; VIDIASOVA; PARKHIMOVICH, 2016). Os segundos benefícios mais citados estão relacionados a aspectos econômicos, como a criação de novas oportunidades de negócios por meio do uso de dados e serviços de dados, além de possibilitar a inovação e a criação de valor (HA; LEE; LEE, 2014).

Outro benefício esperado é a facilidade de consumo e produção de dados usando Ecossistema de Dados. Atores que não possuem as habilidades e recursos para consumir ou fornecer dados podem contratar Prestadores de Serviços ou Intermediários de Dados. Por exemplo, Immonen, Palviainen e Ovaska (2014) apresentam várias funções especializadas relacionadas à provisão de dados. Outro exemplo é Lindman, Kinnari e Rossi (2015), que apresentam as funções de analisar e visualizar dados para ajudar os usuários de dados no processo de consumo de dados. Köster e Suárez (2016) e Schalkwyk, Willmers e McNaughton (2016) enfatizam que os EDs também ajudam a promover a interoperabilidade de dados e serviços, auxiliando, assim, na reutilização e na transparência dos dados.

Outro grande benefício esperado é estimular os atores a interagir e participar. Uma rede bem mantida com atores internos e externos trabalhando coletivamente aumenta as

vantagens competitivas (ZUIDERWIJK et al., 2015). Como afirmado por Immonen, Palviainen e Ovaska (2014), a comunicação entre os atores facilita a prestação de serviços e o compartilhamento de conhecimentos, além de possibilitar diversos tipos de parceria e cooperação. O engajamento e a interação entre os atores também podem ser melhorados com a realização de eventos como *hackathons*, seminários, conferências e competições (KOZNOV et al., 2016).

Dentro os desafios mais citados na literatura destaca-se a falta de conhecimento técnico e de recursos para manter o ecossistema (DONKER; LOENEN, 2017). O segundo desafio mais citado é a alta complexidade para encontrar e consumir os dados. A falta de materiais de referência (SMITH; OFE; SANDBERG, 2016) e conhecimento em como produzir e consumir os dados dificulta o uso e geração de valor a partir dos recursos do ED (ZELETI; OJO, 2016). Outros desafios citados são a falta de participação e interação entre os atores, a falta de uma estrutura organizacional e a falta de privacidade, confidencialidade e confiança.

A falta de participação e interação entre os atores afeta o ecossistema, provocando um menor interesse em participar do ED (KOZNOV et al., 2016), e reduz o compartilhamento de dados, conhecimento e trocas de experiências (SMITH; OFE; SANDBERG, 2016). A falta de estrutura organizacional nesses ecossistemas dificulta que atores internos e externos entendam, implementem e realizem atividades nesses ecossistemas (SHIN, 2016).

Na Web, todos os dias indivíduos criam dados pessoais que são coletados, analisados e usados por entidades públicas e privadas. A violação da privacidade ao publicar os dados e eles serem usados de forma ilegal ou mal intencionada é um grande desafio para os EDs (ZUIDERWIJK et al., 2012). Desafios em torno da privacidade envolvem a invasão de privacidade, falta de segurança, realização de atos ilícitos e o uso inadequado de dados pessoais (SHIN, 2016). Em (MOISO; MINERVA, 2012) os autores destacam a importância e necessidade de se estabelecer padrões e diretrizes técnicos para o gerenciamento da segurança e privacidade.

3 TRABALHOS RELACIONADOS

Até agora, alternativas para medir a saúde do ecossistema de dados ainda são ingênuas, com foco em métricas relativamente simplistas, como número de conjuntos de dados publicados, número e porcentagem de conjuntos de dados baixados, número de conjuntos de dados programados para lançamento, número de APIs e análise básica do site (por exemplo, número de visualizações de páginas, downloads, etc.) (DAWES; VIDIASOVA; PARKHIMOVICH, 2016). A partir do mapeamento sistemático detalhado na Seção 2 não foram encontrados trabalhos que proponham a operacionalização de métricas para avaliar a saúde de EDs. Por isso, fez-se necessário o estudo de como as métricas que avaliam a saúde são operacionalizadas em outras áreas, como em Ecossistemas de Negócios e de *Software*. Somado a isso, encontramos também a operacionalização de métricas em trabalhos na área de Qualidade de *Software*.

Os trabalhos que serão tratados neste Capítulo e que avaliam a saúde dos seus ecossistemas, se inspiraram na área de Ecossistemas Naturais, onde o conceito de saúde de ecossistema foi inicialmente formulado. Costanza (1992) define que um ecossistema natural é saudável se “*ele for estável e sustentável; mantendo sua organização e autonomia ao longo do tempo e sua resiliência ao estresse*”. Somado a isso, Rapport, Costanza e McMichael (1998), se referindo a diversos trabalhos da literatura de Ecossistemas Naturais, define três indicadores para a saúde de ecossistema: *Vigor* que indica o quão ativo ou produtivo um ecossistema é, *Organização* que indica a “*diversidade e o número de interações entre os componentes do sistema*”, e *Resiliência* que indica a habilidade do ecossistema em “*manter a sua estrutura e função na presença de estresse*”. A caracterização de um ecossistema em termos de sua estrutura e função é também discutido por Schaeffer, Hericks e Kerster (1988), na qual a estrutura é identificada como os componentes físicos do ecossistema e a função como as atividades desempenhadas no ecossistema, por exemplo. A estrutura e a função são as dimensões consideradas nos ecossistemas naturais para avaliar qual o impacto que uma exerce sobre a outra. Eles identificam as estruturas como “*número de tipos de organismos, biomassa, entre outros*” e função como “*atividade, produção, decomposição, entre outros*”, os quais são vistos como métricas para avaliar a saúde de ecossistemas.

Nas próximas seções serão detalhados os principais trabalhos que influenciaram na construção do FASED.

3.1 DEN HARTIGH ET AL. 2006

O trabalho apresentado por Hartigh, Tol e Visscher (2006) propõe uma nova definição para a saúde de ECONs e um instrumento com métricas operacionais que pode ser usado

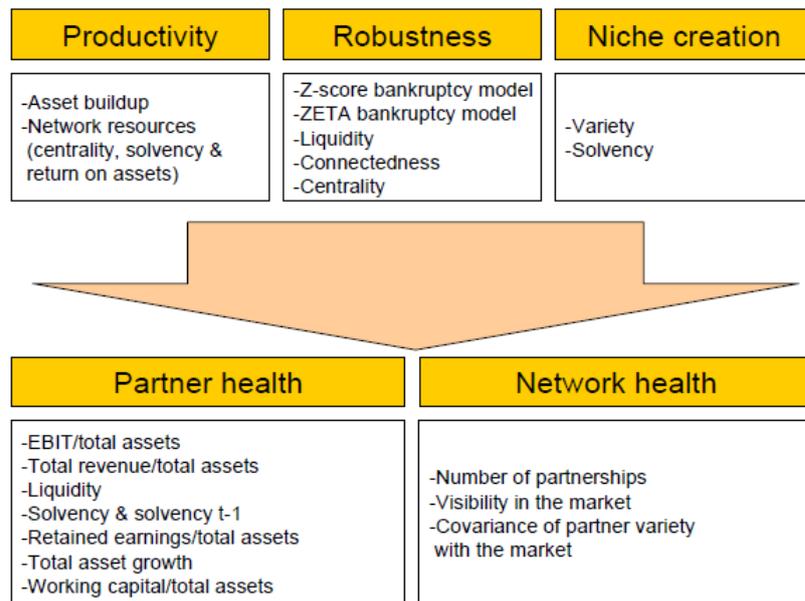
pelos gestores para avaliar a saúde de ECONs. Este trabalho utiliza como referência os determinantes da saúde de ECONs propostos por Iansiti e Levien (2002). Eles definem três determinantes, conhecidos como PRN, para representar a saúde de Ecossistemas de Negócios: a produtividade, a robustez e a criação de nicho. Segue abaixo uma breve explicação sobre cada determinante:

- **Produtividade:** De acordo com Iansiti e Levien (2002), a produtividade é a capacidade de converter matérias-primas de inovação em novos produtos e funções de custo reduzido.
- **Robustez:** De acordo com Iansiti e Levien (2002), a robustez é a capacidade de enfrentar e sobreviver a perturbações e disrupções.
- **Criação de Nicho:** De acordo com Iansiti e Levien (2002), a criação de nicho indica o aumento da diversidade ao longo do tempo por meio da criação de funções que possuam valor.

Apesar de destacarem a importância desses determinantes, Hartigh, Tol e Visscher (2006) propõem converter esses determinantes em novos componentes, os quais avaliam a saúde de um ponto de vista mais gerencial e em múltiplos níveis, como mostrado na Figura 4. Com o intuito de operacionalizar as métricas para avaliar a saúde, Hartigh, Tol e Visscher (2006) definiram a saúde de ECONs como o bem-estar financeiro e a força da rede a longo prazo, e dividiram-na em dois componentes principais: a saúde do parceiro, refletindo o bem-estar financeiro, e a saúde da rede, refletindo a força da rede. A saúde do parceiro avalia a saúde de cada ator do ecossistema. Um ecossistema saudável é composto por atores produtivos que contribuem para a produtividade do ecossistema, enquanto que atores não produtivos enfrentarão dificuldade para sobreviver. Já a saúde da rede é medida em termos de conectividade entre os atores. Atores mais conectados contribuem para a robustez do ecossistema, os quais não são facilmente afetados por choques externos. Somado a isso, um ecossistema saudável contém agrupamentos de diferentes naturezas, que aumenta as chances de criar nichos.

Este trabalho apresenta métricas e processos para avaliar a saúde da rede e dos parceiros. No entanto, a avaliação apenas desses dois componentes não reflete a realidade heterogênea e complexa dos EDs. E os indicadores de sustentabilidade e diversidade não são abordados.

Figura 4 – Relação entre as classificações de Iansiti e Levien (2002) e Hartigh, Tol e Visscher (2006)



Fonte: Hartigh, Tol e Visscher (2006)

3.2 JANSEN 2014

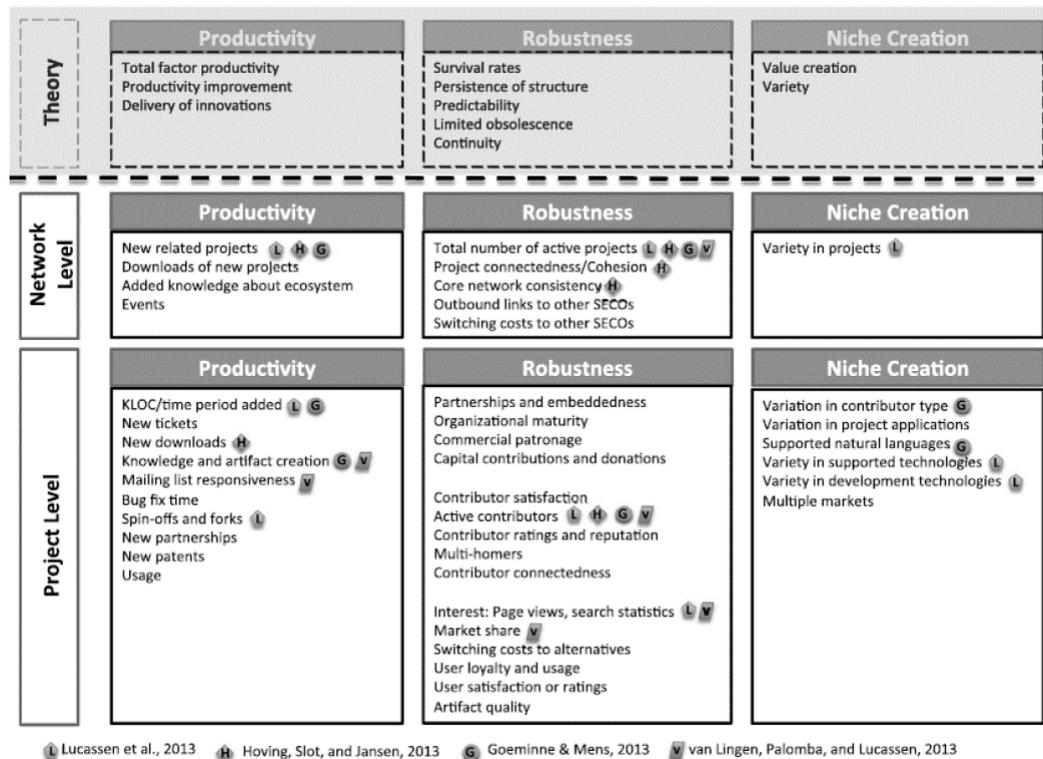
Na literatura de ECOS, a principal fonte de inspiração são os ecossistemas de negócios, pois ambos possuem parâmetros similares de avaliação de saúde. O trabalho apresentado por Jansen (2014) tem o intuito de definir e analisar o funcionamento de ECOS *Open Source*, como base em três indicadores: produtividade, robustez e criação de nicho. O autor sugere que a saúde de um ECOS pode ser definida por dois fatores: longevidade e capacidade de crescimento. Outra observação feita pelo autor é a diferença entre a saúde dos projetos de software e a saúde dos ECOS. A saúde dos projetos de software é avaliada por meio de métricas como: rastreamento e correção de erros, número de *releases* e quantidade de *downloads*. Já a definição da saúde de ECOS depende de fatores como: conexão entre os atores e a capacidade de crescimento do ECOS. O autor apresenta esses fatores como sendo complexos de serem medidos.

A abordagem que o autor utiliza fornece uma visão geral das métricas necessárias para avaliar a saúde de ECOS. Como resultado, uma estrutura chamada *Open Source Ecosystem Health Operationalization* (OSEHO) foi gerada. A Figura 5 apresenta as métricas propostas para avaliar a saúde, agrupando-as por indicador: produtividade, robustez e criação de nicho. As métricas são propostas no contexto teórico e separadas em dois níveis: nível de rede e nível de projeto no contexto de ECOS.

Este trabalho apresenta métricas para avaliar a saúde de ECOS agrupadas pelos indicadores de produtividade, robustez e criação de nicho, mas não aborda os indicadores de sustentabilidade e diversidade. Somado a isso, a estrutura proposta é conceitual e não

detalha como calcular e interpretar os resultados.

Figura 5 – Visão geral do framework OSEHO.



Fonte: Jansen (2014)

3.3 DHUNGANA ET AL. 2010

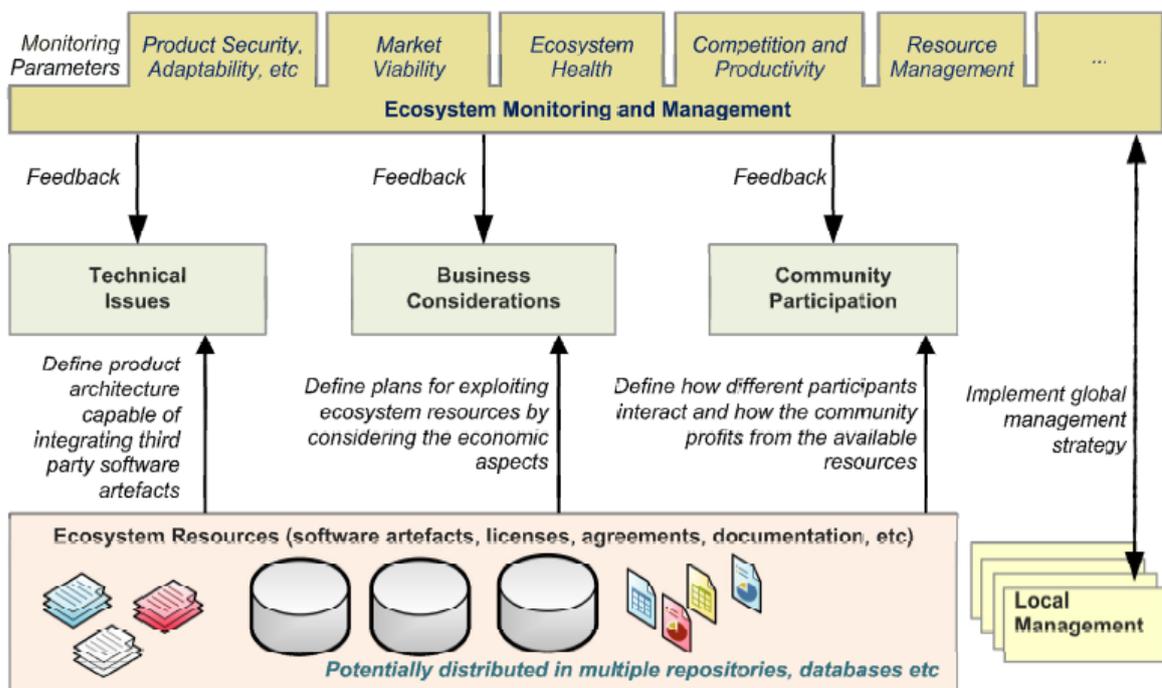
Este trabalho apresenta uma comparação entre Ecossistemas Naturais e de *Software*, e analisa aspectos de sustentabilidade de Ecossistemas Naturais que são relevantes para ECOS. Particularmente, ele foca no desafio de como nutrir fatores de sustentabilidade do ecossistema, pois esse problema é comum nos dois tipos de ecossistemas analisados. A Figura 6 apresenta o *framework* para gerenciar ECOS sustentáveis, o qual é composto pela atividade de definir e monitorar indicadores de saúde. Nesse sentido, Dhungana et al. (2010) afirma que uma comunidade é saudável se ela for sustentável, viável, habitável, igualitária e próspera, e que isso é potencializado quando o ECOS possui atores que se empoderam. Somado a isso, o *framework* também avalia outros aspectos dos ECOS, como gerenciamento de recursos, produtividade, competição, viabilidade de mercado, entre outros.

Ele apresenta as características da sustentabilidade no âmbito dos Ecossistemas Naturais, e relaciona com ECOS. Ele define a sustentabilidade nos ECOS como a capacidade de aumentar ou manter seus atores e sobreviver ao longo do tempo mesmo enfrentando mudanças. Por isso, identificar fatores que fomenta a sustentabilidade do ecossistema envolve a identificação de diferentes aspectos que devem ser cumpridos sem impor controle exces-

sivo. O autor cita algumas métricas da literatura, como número de e-mails, *commits*, *bug fix*, mas não propõe novas formas de medir a sustentabilidade e nem a operacionalização das métricas para avaliar a saúde.

Por fim, ele aborda a questão da diversidade nos Ecosistema Naturais e nos ECOSs. Nos Ecosistemas Naturais a biodiversidade exerce um papel fundamental na resiliência do ecossistema, onde um grupo de espécies pode compensar o outro em momentos de perturbação. Já nos ECOS, a diversidade também se revela crucial, pois envolve diferentes desenvolvedores e grupos de usuários. Por meio de diferentes linguagens de programação, *hardware* e plataformas, possibilita a formação de diferentes grupos de desenvolvimento de artefatos de *software*.

Figura 6 – Framework para gerenciar ECOs sustentáveis.



Fonte: Dhungana et al. (2010)

3.4 FRANCO-BEDOYA ET AL. 2014

De acordo com Franco-Bedoya et al. (2014), “*um ECOS pode ser visto da perspectiva de uma comunidade de atores, organizações e companhias, ou pelo foco técnico e dos aspectos sociais, definindo-o como um grupo de projetos de software e suas comunidades*”. A partir dessas perspectivas, os autores propõem avaliar a qualidade de ECOS definindo características de qualidade que são inerentes ao ECOS.

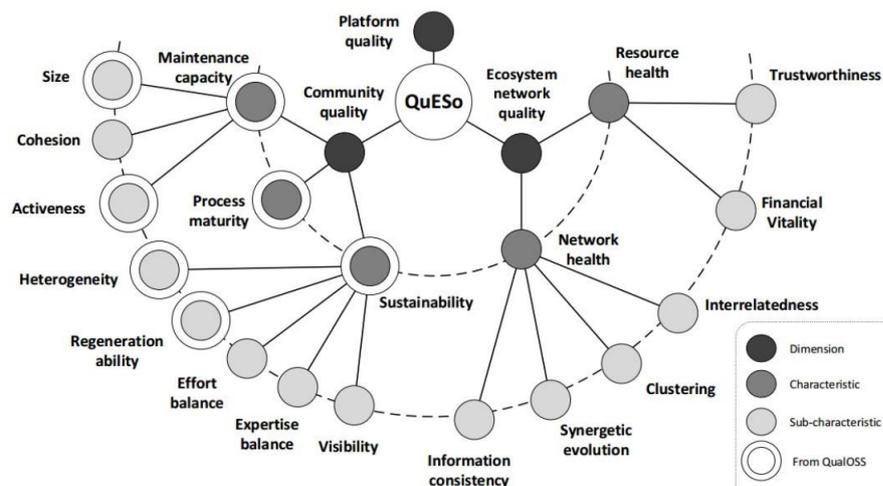
Para elencar essas características, Franco-Bedoya et al. (2014) utilizaram uma Revisão Sistemática da Literatura (RSL). Com as características mapeadas, foi construído o modelo, denominado QuESo. Esse modelo foi composto de dois tipos de elementos: ca-

racterísticas de qualidade e métricas. As características de qualidade correspondem aos atributos de software relevantes para a avaliação, organizadas em uma hierarquia composta no nível mais alto por 3 dimensões. As dimensões são a plataforma de ECOS, a comunidade e sua rede de atores. As dimensões são divididas em características e estas, por sua vez, são divididas em sub-características. Algumas características foram extraídas do modelo de qualidade para softwares *Open Source*, denominado QualOSS (SOTO; CIOLKOWSKI, 2009). A Figura 7 ilustra o modelo de qualidade de software QuESo.

Para exemplificar as métricas que compõem o modelo pode-se citar a característica Capacidade de Manutenção, a qual é composta pela sub-característica Tamanho, que por sua vez pode ser medida pelo número de atores ativos, investidores, entre outros, número de erros corrigidos, dentre outros.

O QuESo mostra-se capaz de analisar a qualidade de ECOSs, entretanto ele não aborda a influência da qualidade da plataforma na qualidade dos produtos, presentes no ECOS. Somado a isso, o modelo e as métricas apresentadas, assim como seus indicadores e características, são apresentados de forma conceitual, não possuindo nenhuma forma de cálculo e interpretação.

Figura 7 – Modelo de qualidade QuESo.



Fonte: Franco-Bedoya et al. (2014)

3.5 CARVALHO ET AL. 2017

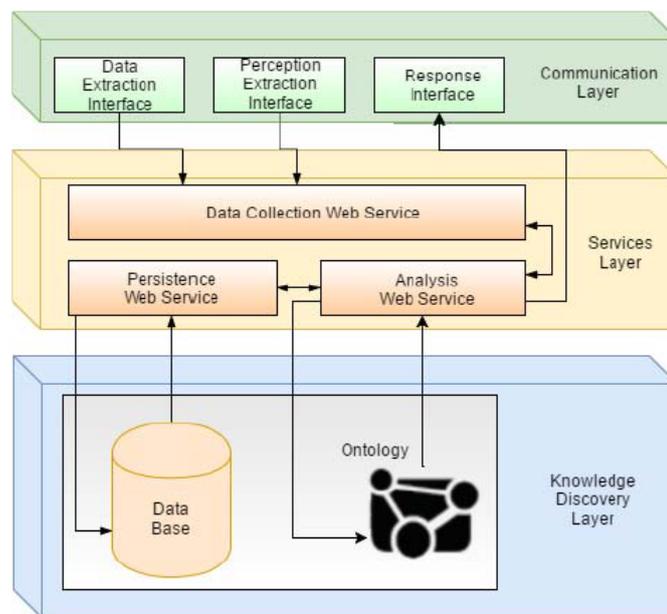
Este trabalho propõe uma arquitetura de três camadas com o objetivo de avaliar a saúde de ECOS, denominado HEAL ME. A arquitetura é composta pelas camadas de comunicação, serviço e descoberta de conhecimento, ver Figura 8. A camada de comunicação é responsável pelo acoplamento da arquitetura à estrutura do ECOS, bem a suas aplicações. Essa camada captura os dados dos componentes (*e.g.* dados de projetos, comunidade, usuários, desenvolvedores, plataforma e seus produtos) do ECOS de forma automática.

Ela também provê uma interface para a inclusão manual de dados não disponíveis pelas APIs.

A camada de serviço é responsável pela captura dos dados oriundos da interface. Essa camada insere os dados nas ontologias, as quais são responsáveis pela interpretação e depois devolve os dados para a camada de comunicação para que os dados sejam processados. Este trabalho apresenta uma ontologia chamada OntoHealth, que modela o domínio de ECOSs. Por fim, a camada de descoberta de conhecimento contém a ontologia e as regras de avaliação. Essa camada contém também o repositório de dados processados pela ontologia e dados sobre a saúde do ECOS.

Com isso, como forma de avaliar a saúde, o HEAL ME utiliza os indicadores propostos nos trabalhos de Dhungana et al. (2010) e Iansiti e Levien (2002) para medir a saúde dos ECOSs. Além dos trabalhos anteriores, os autores também utilizaram o trabalho (JANSEN, 2014) e o modelo de qualidade proposto por Franco-Bedoya et al. (2014) como fontes para extrair as métricas. Carvalho et al. (2017a) estruturou a avaliação da saúde em uma hierarquia de indicadores, características e métricas, onde podem ser encontrados 5 indicadores, 9 características e 58 métricas. E, por fim, um estudo de caso foi feito com o intuito de avaliar a viabilidade e aplicabilidade do HEAL ME, no contexto de um ECOS científico.

Figura 8 – Arquitetura HEAL ME.



Fonte: Carvalho et al. (2017a)

3.6 SÍNTESE DO CAPÍTULO

Este Capítulo apresentou modelos e métricas que influenciaram na construção do FASED. Esses trabalhos apresentam a importância de avaliar a saúde e o impacto dessa avaliação

no funcionamento dos componentes do ecossistema. Para alcançar esse objetivo foram propostos indicadores que avaliam a saúde de ecossistemas, como Produtividade, Robustez e Criação de Nicho para os ECONs e a Sustentabilidade e Diversidade nos ECOSs. Somado a isso, vimos o modelo de qualidade QuESo que propõe características para avaliar a qualidade de ECOS e a arquitetura HEAL ME com o intuito de avaliar a saúde de ECOS reunindo os 5 indicadores propostos nos trabalhos de ECONs e ECOSs anteriores. No entanto, nenhum desses trabalhos aborda o domínio de EDs e nem todas as métricas propostas se adequam à realidade dos EDs. Portanto, o próximo Capítulo descreverá o FASED, com seus indicadores, características e métricas detalhadas para avaliar a saúde de EDs.

4 FASED: FRAMEWORK PARA AVALIAÇÃO DA SAÚDE DE ECOSISTEMAS DE DADOS

Como mostrado no Capítulo 3, em sua maioria, os *frameworks* que avaliam a saúde de ecossistemas são específicos para *software* e negócios. Com isso, analisando a organização, os atores e papéis, as relações entre componentes e as diferentes complexidades dos EDs, concluímos que os *frameworks* propostos em outros domínios não se adéquam à realidade dos EDs. Somado a isso, como identificado por Oliveira, Lima e Lóscio (2019), a formalização de métricas específicas para avaliar a saúde de EDs é um tópico ainda não explorado na literatura.

Os EDs possuem, dependendo da complexidade do Ecossistema, um crescimento rápido na quantidade de dados disponibilizados e consumidos direta ou indiretamente pelos atores, alavancando assim a produção de novos recursos e revelando a necessidade de se ter ambientes saudáveis para lidar com essa nova dinâmica. Portanto, diante dessa necessidade, propomos neste trabalho um *framework* para avaliar a saúde desses Ecossistemas utilizando métricas que revelem o panorama da saúde dos EDs.

Um *framework* pode ser definido como uma estrutura de apoio ou inclusão de outro artefato, especialmente um suporte esquelético usado como base para algo que está sendo construído (Oxford Dictionary, 2019). Segundo Jabareen (2009), *frameworks* conceituais são constituídos por um conjunto de conceitos interligados que apoiam o entendimento de um fenômeno do mundo real. O *framework* de avaliação deste trabalho possui um caráter conceitual, pois utiliza uma linguagem notacional para fornecer uma representação clara de um conjunto de elementos e das relações existentes entre eles. Os elementos, *e.g.* indicadores, características, atributos e métricas, fazem parte da área de conhecimento de saúde de ecossistemas e modelos de qualidade, e ajudam a explicar o domínio em questão (NOY; MCGUINNESS et al., 2001).

O *framework* conceitual proposto neste trabalho tem o objetivo de apoiar os atores do ED a tomarem melhores decisões baseados na avaliação da saúde do ecossistema. De uma forma geral, o *framework* proposto contribuirá para: (i) possibilitar que os EDs se tornem sustentáveis e autorreguláveis por meio da identificação de problemas que estejam afetando a sua saúde, (ii) identificar nichos em diferentes domínios, (iii) identificar mudanças significativas na produção e consumo de recursos, (iv) mapear o histórico de resultados ao longo do tempo e (v) mapear características que estejam ausentes no ecossistema mas que são de suma importância para que ele se mantenha saudável.

Este trabalho é o primeiro passo na área de avaliação da saúde de Ecossistemas de Dados e tem como objetivo propor e explorar métricas genéricas baseadas nos componentes de EDs e nas atividades relacionadas para possibilitar a avaliação do ecossistema em estudo. Analisando os domínios dos EDs identificados no mapeamento sistemático,

percebe-se a presença de relacionamentos e recursos heterogêneos (OLIVEIRA; LÓSCIO, 2018), o que possibilita um levantamento abrangente de métricas e, conseqüentemente, a avaliação se torna mais complexa.

4.1 METODOLOGIA PARA CONSTRUÇÃO DO FRAMEWORK

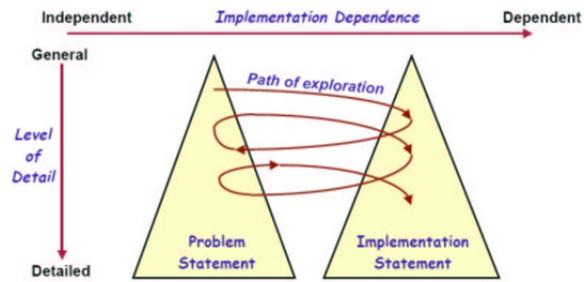
A construção do *framework* se deu por um processo iterativo e incremental. Como resultado do mapeamento sistemático apresentado na Seção 2, uma base teórica sobre Ecossistemas de Dados foi consolidada. Somado a isso, uma base teórica sobre *frameworks* de avaliação da saúde e modelos de qualidade foi construída pelo estudo *ad-hoc* detalhado na Seção 3. O estudo dos *frameworks* de avaliação da saúde de ecossistemas de negócio e de software, como também os modelos de qualidade de avaliação de software, nos ajudou a entender como eles foram construídos, desde o estudo do contexto do domínio até a avaliação, utilizando métricas específicas para cada indicador. Ambos os estudos guiaram a construção do FASED.

O *framework* recebe influências da ISO/IEC 25000:2014 e de alguns trabalhos identificados durante o estudo *ad-hoc*. A norma ISO/IEC 25000:2014 (*International Organization for Standardization and International Electrotechnical Commission 25000:2014*), a qual apresenta o SQuaRE (*Systems and software Quality Requirements and Evaluation*), reúne uma série de padrões internacionais em métricas e modelos de qualidade, como também em avaliação e requisitos de qualidade para sistemas e produtos de software. A escolha dessa norma se deu em função da sua natureza genérica, por conter conceitos de qualidade em alto nível, possibilitar a construção de hierarquias de características de qualidade e por ser um padrão amplamente difundido na literatura. Dentre os trabalhos que influenciaram o *framework*, estão os trabalhos de Jansen (2014) e Carvalho et al. (2017a) na área de Ecossistema de *Software*, e o trabalho de Hartigh, Tol e Visscher (2006) na área de Ecossistemas de Negócios. Esses trabalhos propõem *frameworks* de avaliação da saúde utilizando uma abordagem *top-down* que identifica indicadores de saúde como ponto de partida e refina até identificar as métricas.

Do conjunto inicial de conceitos, o *design* do FASED foi continuamente verificado e refinado. O processo de construção, no qual as fases de refinamento e verificação foram executadas ciclicamente, é inspirado no modelo de *Twin Peaks* (NUSEIBEH, 2001). O processo é iterativo e produz progressivamente requisitos e especificações mais detalhadas, como mostra a Figura 9.

Sendo assim, a construção começou a partir da elaboração de uma versão inicial do *framework* baseada no estudo *ad-hoc* de *frameworks* propostos em outros tipos de ecossistemas e nos modelos de qualidade de avaliação de *software*. Essa versão inicial foi melhorada por meio de ciclos de verificação e refinamento. Todos os elementos que compõem o FASED foram verificados quanto a contribuição de cada um para a definição do

Figura 9 – Modelo Twin Peaks.



Fonte: Nuseibeh (2001)

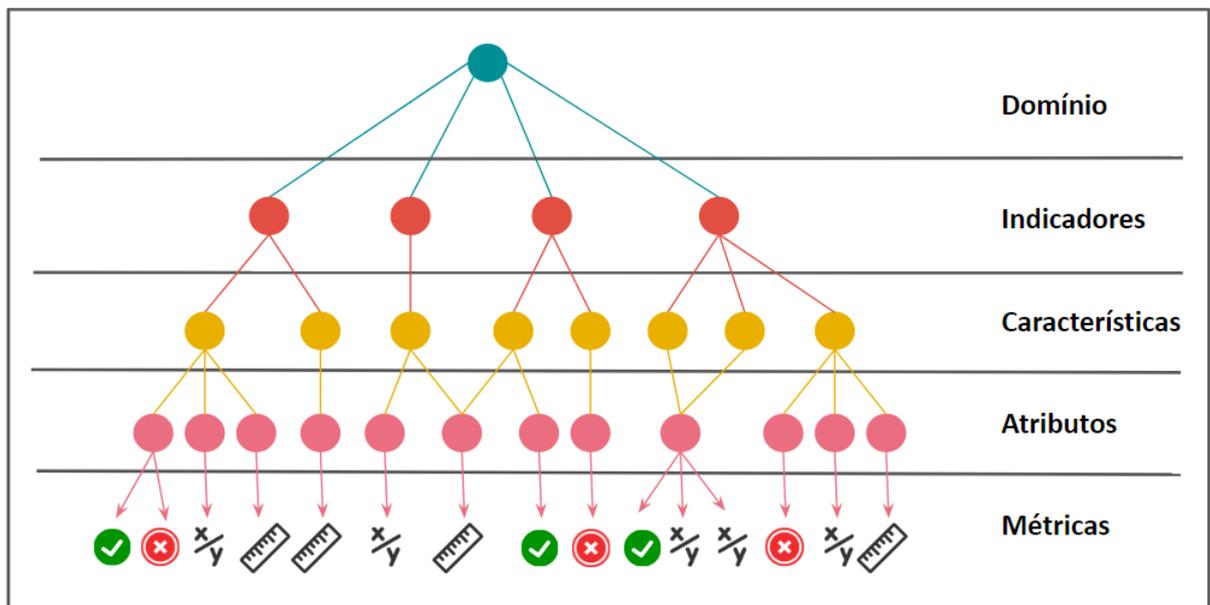
framework e para a avaliação da saúde de EDs. Após isso, os resultados do processo de verificação foram usados no refinamento do *framework* proposto.

4.2 FASED: FRAMEWORK PARA AVALIAÇÃO DA SAÚDE DE ECOSISTEMAS DE DADOS

O FASED é estruturado para guiar a avaliação da saúde de Ecosistemas de Dados de forma prática. Ele é baseado em um conjunto de indicadores, características, atributos e métricas. O FASED é embasado no corpo de conhecimento de Ecosistemas de Dados, *frameworks* para avaliação da saúde de ecossistemas e modelos de qualidade de avaliação de *software*, o qual foi adotado como referência para a construção do *framework*. Por esse ser o primeiro trabalho na área, nosso objetivo é propor um *framework* de caráter genérico que pode ser adaptado para o contexto específico do ED no qual for aplicado. Somado a isso, é importante ressaltar que o *framework* proposto nesta dissertação estende e adapta conceitos oriundos da pesquisa em Ecosistemas de Software e Ecosistemas de Negócios.

O FASED segue uma abordagem descritiva, pois sugere quais indicadores, características, atributos e métricas podem ser considerados pelos atores que vão executar a avaliação da saúde dos seus EDs. A utilização da abordagem descritiva permite a adaptação do *framework* à realidade do ED a ser avaliado. Essa abordagem é interessante, pois a identificação de aspectos específicos do ecossistema em estudo vai aumentando ao passo que os elementos do *framework* vão sendo mapeados.

Figura 10 – Estrutura do framework



Fonte: A autora (2019)

A Figura 10 ilustra a estrutura do FASED, tomando como ponto de partida a análise do domínio do ED, passando pelo mapeamento dos indicadores e refina até chegar no nível mais baixo da hierarquia que são as métricas. Cada nível é um elemento que cons-

titui o *framework*. Os conceitos de "indicador", "característica" e "métrica" são baseados na ISO/IEC 25000:2014 e no trabalho de (SOTO; CIOLKOWSKI, 2009), e adaptados para o contexto de EDs. Cada conceito está descrito abaixo:

- **Indicador:** corresponde à interpretação dos resultados de um conjunto de métricas quantitativas e/ou qualitativas que influenciam nas tomadas de decisão do ecossistema.
- **Característica:** corresponde a um aspecto abstrato relacionado a um ou mais atributos do ED que tem o objetivo de agrupar métricas relacionadas à mesma característica.
- **Atributo:** corresponde aos constructos principais de EDs, *e.g.* atores, recursos, papéis e relações.
- **Métrica:** corresponde a um aspecto concreto que podemos medir no ED e está relacionada a um atributo e a uma característica.

Nas próximas seções serão apresentados com mais detalhes os indicadores, características, atributos e métricas que compõem o FASED.

4.2.1 Indicadores

A estrutura do *framework* detalhada anteriormente tem como ponto de partida os indicadores. A escolha dos indicadores foi uma fase importante desse trabalho, pois como não há trabalhos para avaliar a saúde de EDs, nós estudamos como essa avaliação tem sido feita em outros tipos de ecossistemas. Encontramos *frameworks* que avaliam a saúde de ecossistemas nas áreas de *software* e negócios e, a partir daí, definimos quais indicadores se adequam e influenciam a saúde dos EDs. Ver mais detalhes sobre EDs na Seção 2.

O uso de indicadores possibilita a agregação de características que podem ser avaliadas, por meio de métricas, com o objetivo de obter insumos que podem ser prontamente utilizados em tomadas de decisões no ED. A interpretação dos resultados obtidos por meio dos indicadores tem o objetivo de mostrar para os *stakeholders* quais aspectos do ecossistema, *e.g.* uso de um aplicativo ou investimento governamental, estão saudáveis e quais não estão. Somado a isso, os resultados da avaliação podem influenciar na dinâmica do ecossistema e no gerenciamento de riscos e problemas.

Para o mapeamento de quais indicadores poderiam ser adequados para a avaliação da saúde de EDs, nós pesquisamos os indicadores propostos nos *frameworks* de avaliação de saúde para os ecossistemas de negócios e de *software*. Os indicadores mais utilizados na literatura são os indicadores de Produtividade, Robustez e Criação de Nicho. Esses indicadores são a base de *frameworks* de avaliação de saúde, amplamente difundidos na literatura, e são os indicadores propostos por Iansiti e Levien (2004b) para avaliar a saúde

de ecossistemas de negócios. Após a publicação desse trabalho, a maioria dos *frameworks* de avaliação de saúde, tanto do domínio de negócios quanto de software, também adotaram esses indicadores, por exemplo (CARVALHO et al., 2017a) (JANSEN, 2014) (MANIKAS; HANSEN, 2013a) (HARTIGH; TOL; VISSCHER, 2006). Contudo, outro indicador também aparece na avaliação da saúde de ecossistemas naturais e de software, o indicador de Sustentabilidade, o qual é discutido no trabalho de Dhungana et al. (2010) e proposto por Carvalho et al. (2017a) como um novo indicador para avaliação da saúde de ecossistemas de software.

Após esse estudo, escolhemos os indicadores que se aplicam à realidade dos EDs e que podem indicar o panorama da saúde dos mesmos. A seguir descrevemos como esses indicadores são utilizados em outras áreas e como eles serão aplicados na avaliação da saúde de EDs:

- **Produtividade:** De acordo com Iansiti e Levien (2002), o indicador de produtividade nos Ecossistemas da Natureza é definido como a capacidade efetiva de converter matérias-primas em organismos vivos. Seguindo essa definição e considerando que os Ecossistemas de Negócios são sujeitos constantemente a novas condições, como novas tecnologias, processos e demandas, Iansiti e Levien (2002) definem a produtividade como a capacidade de converter matérias-primas de inovação em novos produtos e funções de custo reduzido. Tomando como ponto de partida as definições apresentadas nos contextos de natureza e negócios, definimos *a produtividade de EDs como a capacidade de produzir novos recursos (e.g. dados, soluções, serviços), permitir o consumo desses recursos e realizar atividades que promovam a produtividade*. Para avaliar o indicador de produtividade em EDs se faz necessário o uso de métricas que meçam a capacidade de produção de novos recursos, a partir do volume produzido; a capacidade de consumo, por meio dos números de acessos e downloads de dados e soluções; e a existência de atividades que promovam a produtividade, como eventos e disponibilização de materiais de referência que facilitem o uso de recursos.
- **Robustez:** De acordo com Iansiti e Levien (2002), o indicador de robustez nos Ecossistemas da Natureza é definido como a capacidade de persistir diante de mudanças no meio ambiente. Similar a essa definição, Iansiti e Levien (2002) definem que os Ecossistemas de Negócio deveriam ser capazes de enfrentar e sobreviver a perturbações e disrupções. Similar às definições apresentadas, *um ED pode ser considerado robusto se ele possuir a capacidade de se manter estável ao enfrentar disrupções*. Neste trabalho o termo disrupção significa perturbação ou problemas que interrompam um evento, atividade ou processo (Oxford Dictionary, 2019). Dito isso, o indicador de robustez pode ser avaliado por métricas que meçam a sua consistência financeira, a gestão e o estado dos recursos perante disrupções.

- **Criação de Nicho:** De acordo com Iansiti e Levien (2002), o indicador de criação de nicho em Ecossistemas da Natureza exibe variedade e suporta a diversidade de espécies. Similarmente, Iansiti e Levien (2002) define que esse indicador em Ecossistemas de Negócio indica o aumento da diversidade ao longo do tempo por meio da criação de funções que possuam valor. Então, levando em consideração essas definições, *no contexto de EDs esse indicador também exibe a variedade dos elementos do ecossistema com o objetivo de identificar oportunidades para o surgimento de novos nichos*. Esse indicador identifica a existência e variedade de atores, papéis e recursos.
- **Sustentabilidade:** De acordo com III, Torn e Tateno (1996), sustentabilidade é um dos principais desafios em qualquer ecossistema. Um ecossistema natural sustentável mantém a diversidade dos principais grupos funcionais, a produtividade e os ciclos bio-químicos, mesmo frente a eventos que perturbem o estado natural desse ecossistema. Similar à definição de III, Torn e Tateno (1996), Dhungana et al. (2010) considera um ecossistema de *software* sustentável quando ele aumenta ou mantém seus produtos, recursos, membros e relacionamentos por longos períodos de tempo e consegue sobreviver às mudanças como novas tecnologias, novos produtos, competidores. Em resumo, sustentabilidade é a capacidade de manter os produtos ou serviços que o Ecossistema de *Software* desenvolve por um longo período de tempo. Com isso, similar às definições apresentadas nos contextos de natureza e *software*, consideramos que *os EDs são sustentáveis quando possuem a capacidade de aumentar ou manter seus recursos, atores e relacionamentos por longos períodos de tempo e sobreviver às mudanças, como novas tecnologias, entrada e saída de atores, competidores de mercado*. Para avaliar o indicador de sustentabilidade em EDs se faz necessário o uso de métricas que meçam a capacidade de crescimento, o equilíbrio de esforços entre os atores, engajamento dos atores, visibilidade dos recursos e do próprio ED, retorno de insumos, e.g. *feedback*, como também a qualidade dos dados disponíveis.

Por causa da realidade distribuída e heterogênea dos EDs (OLIVEIRA; LÓSCIO, 2018), os indicadores propostos acima são de caráter generalista e devem ser adaptados ao domínio do ED avaliado. Ver cenário motivacional na Seção 5.2 que aplica esses indicadores.

4.2.2 Características

Cada indicador contém uma ou mais características, as quais representam aspectos abstratos relacionados aos principais constructos dos EDs. Durante a construção do *framework* proposto, a etapa de definição das características se deu após o mapeamento dos indicadores. Essa etapa foi baseada no estudo dos aspectos relacionados a cada indicador, dos principais atributos (e.g. dados, soluções) dos EDs e no estudo das características

propostas na literatura. Identificamos trabalhos no contexto de modelos de qualidade que utilizam características como forma de representar aspectos abstratos relacionados ao indicador que se quer medir e que são referenciados pelos *frameworks* de avaliação da saúde de ecossistemas de software (CARVALHO et al., 2017a) (CARVALHO et al., 2017b). Além das características encontradas na literatura, identificamos características específicas para o contexto de ED que se aplicam ao *framework* proposto neste trabalho.

Levando em consideração os indicadores apresentados na Seção 4.2.1, apresentamos as seguintes características:

1. Produtividade

- **Produção:** indica o estado atual da produção de recursos no ED. O objetivo é medir o volume de recursos produzidos, pois a quantificação demonstra a capacidade produtiva do ecossistema e, por consequência, agrega valor ao ecossistema.
- **Consumo:** indica o estado atual do consumo de recursos do ED. O objetivo é medir o volume de consumo dos recursos produzidos pelo ecossistema, pois a quantificação desse consumo estimula a manutenção dos recursos existentes e a produção de novas soluções.
- **Atividade:** indica a existência de atividades (*e.g. hackathons, workshops, materiais de referência*) que promovem a produção e consumo de recursos pelos atores do ED.

2. Robustez

- **Consistência Financeira:** indica a habilidade do ED em se manter com os investimentos recebidos e recursos financeiros produzidos por ele, enquanto atrai novos investidores e parceiros (BEDOYA et al., 2016).
- **Garantia de Disponibilidade:** indica se existe mecanismos que garantam a disponibilidade contínua dos recursos (*i.e.* dados, aplicações, serviços, infraestrutura) mesmo em situações de instabilidade.
- **Gestão de Riscos:** indica a existência de ações e recursos que auxiliam na gestão do ecossistema, como auditoria, estruturas organizacionais bem definidas, atores e processos. Eles auxiliam na prevenção de riscos caso aconteça perturbações no ecossistema.

3. Criação de Nicho

- **Variedade:** indica a existência e a capacidade de promover a criação de nichos analisando a variedade dos elementos do ecossistema, como atores, papéis e recursos.

4. Sustentabilidade

- **Habilidade de Regeneração:** indica a evolução da rede de atores, por meio do número de atores que entram e interrompem sua participação no ecossistema. Essa análise mostra se o ED já precisou se regenerar no passado, indicando que essa característica é presente e o ecossistema tem mais chances de sobreviver no futuro caso haja perdas significativas de atores (BEDOYA et al., 2016).
- **Equilíbrio de Esforços:** indica a distribuição de esforços no ED, pois quanto maior a concentração do conhecimento em pequenos grupos, maior o risco se esses grupos não estiverem presentes no ecossistema (BEDOYA et al., 2016).
- **Heterogeneidade:** indica a diversidade do ED que influencia no seu crescimento e sobrevivência (BEDOYA et al., 2016). Essa característica pode ser identificada pela diversidade na distribuição geográfica do ED.
- **Engajamento:** fornece informações sobre a participação dos atores no ED. A participação dos atores pode acontecer por meio de eventos que promovam o ED, publicações de recursos que aumentem a visibilidade do ecossistema ou pelo uso de canais de comunicação por meio de *feedback*. O aumento do engajamento mostra o interesse dos atores no ecossistema e atrai novos colaboradores para contribuir e dar suporte ao ecossistema.
- **Qualidade:** indica se os dados estão aptos para uso pelos consumidores (STRONG; LEE; WANG, 1997). Essa característica reúne informações sobre a qualidade dos recursos disponibilizados pelo ecossistema, avaliando acessibilidade, consistência, confiabilidade, influenciando assim o consumo e, conseqüentemente, a sobrevivência do ecossistema.

Por causa da realidade distribuída e heterogênea dos EDs (OLIVEIRA; LÓSCIO, 2018), as características propostas acima são de caráter generalista e devem ser adaptadas ao domínio do ED avaliado. Ver cenário motivacional na Seção 5.2 que aplica essas características.

4.2.3 Atributos

Os atributos selecionados são os constructos principais de um Ecossistema de Dados, que apesar de representarem os principais elementos dos EDs, eles são dinâmicos e podem evoluir com o tempo (OLIVEIRA et al., 2018). Os atributos são:

- **Ator:** entidade autônoma, como empresa, instituição ou indivíduo, que desempenha um ou mais papéis específicos em um ED. Um ator é considerado um elemento básico do ED. Um conjunto de interesses motiva os atores e cada um tem expectativas e

capacidades diferentes. Os atores geralmente se engajam no ecossistema, os quais podem ter incentivos para permanecer ativos no ecossistema.

- **Papel:** função desempenhada por um ator do ED. Está relacionado a um conjunto de deveres e atividades. Diversos papéis podem ser identificados nos EDs, os quais são responsáveis por diferentes deveres e atividades. No mínimo, são identificados os papéis de consumidor e produtor nos EDs.
- **Relacionamento:** interação entre os atores do ED. Os relacionamentos são geralmente baseados em interesses comuns ou são relacionados ao papel desempenhado pelo ator. Os relacionamentos variam de acordo com diversos aspectos, como econômicos, políticos, culturais e tecnológicos. Os atores trocam dados ou outros tipos de recursos pelas transações dos relacionamentos, os quais produzem valor e envolvem custos.
- **Recurso:** produto ou capacidade útil ou de valor, produzido, fornecido, curado ou consumido pelos atores. Em EDs, os recursos variam de conjuntos de dados e software baseado em dados, até infraestrutura e serviços.

4.2.4 Métricas

O uso de métricas, para medir aspectos concretos relacionados aos atributos do ED, fornece informações relevantes para os atores envolvidos no ED. Os resultados obtidos por meio da aplicação das métricas auxiliam no gerenciamento de riscos, detecção de problemas e geram um panorama do estado da saúde do ecossistema. No *framework* proposto neste trabalho há algumas métricas extraídas de *frameworks* que avaliam a saúde de ecossistemas de software, de modelos de qualidade de *software* e outras métricas que foram criadas pelo autor para suprir necessidades específicas do contexto de EDs.

Nas próximas Seções 4.2.4.1, 4.2.4.2, 4.2.4.3 e 4.2.4.4 serão apresentadas as métricas resultantes do estudo. Cada métrica foi detalhada e seu indicador de saúde foi apontado, assim como a característica a qual ela pertence. Vale destacar que a maioria das métricas referenciadas da literatura, exibem apenas nome e descrição, sem detalhes e formalização.

Para facilitar o entendimento, as métricas foram agrupadas por indicadores e características. Cada métrica possui as seguintes informações:

1. Indicador: ao qual a métrica pertence;
2. Nome da métrica;
3. Descrição: objetivo da métrica;
4. Medida e Fórmula: mostra como a métrica pode ser calculada;
5. Interpretação: mostra como o resultado da métrica pode ser interpretado;

6. Unidade: unidade de medida do resultado;
7. Atributo: constructo do ecossistema relacionado com a métrica;
8. Referência Bibliográfica: métrica criada pelo autor ou já referenciada por outros trabalhos.

Com o detalhamento de cada métrica, provemos um direcionamento de como aplicá-la em um cenário real e como interpretar os resultados. Para facilitar a obtenção dos dados a serem analisados, as métricas possuem atributos vinculados. Por meio da aplicação do *framework* e de posse dos resultados coletados, é possível identificar a presença dos indicadores ou indícios de problemas que podem estar influenciando a saúde de EDs. Com esses resultados, conseguimos definir o panorama da saúde do ecossistema em estudo e revelar quais características precisam de mais atenção.

Tendo em vista o caráter generalista das métricas, apresentamos alguns sinônimos para os termos utilizados nas métricas que devem ser adaptados de acordo com o domínio do ED que se deseja avaliar a saúde:

- Recursos: dados, soluções, serviços, infraestrutura, conjuntos de dados, metadados, entre outros;
- Soluções: aplicações *Web*, aplicativos móveis, serviços, plataforma, portal, infraestrutura, entre outros;
- Eventos: palestras, *workshops*, *hackathons*, entre outros;
- Materiais de referência: tutoriais, relatórios, mapas mentais, fluxogramas, entre outros;
- Literatura: artigos, patentes, normas, capítulos, dissertações, teses, entre outros.

De posse dos termos adequados para o ED em estudo, deriva-se as medidas e fórmulas para calcular as métricas. O resultado após a aplicação da fórmula fica armazenado na variável X , a qual posteriormente será interpretada. Os resultados podem ser: números inteiros, razão entre inteiros e valor binário (*e.g.* sim ou não). Os números inteiros podem ser valores exatos (*e.g.* métricas C1, P8, entre outras) ou calculados utilizando o somatório que fornece um panorama da métrica em relação a um período de tempo desejado (*e.g.* métricas P1, S2, R13, entre outras). A fórmula desse tipo de métrica inclui duas variáveis: T e N_T , onde T é o número que representa o período de tempo (*e.g.* número de horas, dias, meses, anos) e N é o valor medido em relação à variável de tempo T . Essas métricas podem ser aplicadas mais de uma vez na mesma avaliação alterando o período de tempo, pois assim permite a comparação entre resultados da mesma métrica em diferentes períodos.

Somado a isso, as fórmulas que envolvem razão entre inteiros (*e.g.* P6, R11, entre outras) fornecem um valor decimal que pode ser convertido em porcentagem e assim

facilitar a interpretação. Essas métricas fornecem dois tipos de informações: a porcentagem do que se quer medir e a porcentagem do complemento, as quais nos ajudam a ter uma visão mais abrangente da situação real da métrica. Por fim, as métricas que resultam em valor binário (e.g. S11, R14 entre outras) revelam a presença ou ausência do que se quer medir.

Com o resultado de cada métrica, o valor da variável X , é possível fazer uma comparação com o valor do parâmetro P . Esse parâmetro pode ser uma meta definida pelos próprios gestores do ED, pode ser um valor que foi medido em uma avaliação anterior do *framework* ou em outro período de tempo. Os valores utilizados como parâmetro podem ser alterados com o tempo e servem como referência para que estratégias e medidas possam ser tomadas após a interpretação dos resultados.

As explicações acima sobre as fórmulas, interpretações e sinônimos de termos apresentados nesta Seção são aplicáveis a todas as métricas propostas no FASED.

Na Seção 5.2 é apresentado um cenário motivacional com a aplicação das métricas propostas neste trabalho.

4.2.4.1 Métricas de Produtividade

As métricas de produtividade são agrupadas por características. A divisão por características facilita o entendimento do que se quer medir. Dentre as características de produtividade tem-se: Produção, Consumo e Atividade. Essas três características agrupam métricas que calculam dois aspectos diferentes que influenciam na saúde: o estado atual da produtividade e a promoção da produtividade no ED.

O estado atual da produtividade é calculado por métricas que avaliam o volume de recursos produzidos e os números associados ao consumo desses recursos. Já as métricas que medem a promoção da produtividade focam na presença de eventos que promovem os recursos e na existência de materiais de referência que auxiliam na produção e consumo de recursos do ED.

Tabela 4 – Métricas do Indicador de Produtividade para Avaliar a Saúde de EDs

ID	Indicador	Métrica	Descrição	Medida e Fórmula	Interpretação	Unidade	Atributo	Referência Bibliográfica
Indicador: Produtividade - Característica: Produção								
P1	Produtividade	Volume de dados produzidos	Volume de dados produzidos pelo ED e disponíveis para uso	$X = \sum_1^T N_T$ $N_T = \text{Volume de dados}$ $T = \text{Período de tempo}$	$X \geq P$ P = Parâmetro Quanto maior o volume de dados produzidos melhor	Un - Unidade	Recursos	Autor (2019)
P2	Produtividade	Volume de soluções produzidas	Volume de soluções produzidas pelo ED e disponíveis para uso	$X = \sum_1^T N_T$ $N_T = \text{Volume de soluções}$ $T = \text{Período de tempo}$	$X \geq P$ P = Parâmetro Quanto maior o volume de soluções produzidas melhor	Un - Unidade	Recursos	Autor (2019)
Indicador: Produtividade - Característica: Consumo								
P3	Produtividade	Número de acessos	Número de acessos realizados aos recursos	$X = \sum_1^T N_T$ $N_T = \text{Número de acessos}$ $T = \text{Período de Tempo}$	$X \geq P$ P = Parâmetro Quanto maior o número de acessos realizados melhor	Un - Unidade	Recursos	Autor (2019)
P4	Produtividade	Número de downloads	Número de downloads realizados nos recursos	$X = \sum_1^T N_T$ $N_T = \text{Número de downloads}$ $T = \text{Período de Tempo}$	$X \geq P$ P = Parâmetro Quanto maior o número de downloads realizados melhor	Un - Unidade	Recursos	Jansen (2014) Franco Bedoya et al. (2016)
P5	Produtividade	Número de soluções externas que consomem dados do ED	Número de soluções de ecossistemas externos que consomem os dados do ED	$X = \sum_1^T N_T$ $N_T = \text{Número de soluções de ecossistemas externos}$ $T = \text{Período de Tempo}$	$X \geq P$ P = Parâmetro Quanto maior o número de soluções melhor	Un - Unidade	Recursos	Autor (2019)

P6	Produtividade	Soluções do ED que consomem os próprios dados	Razão entre a quantidade de soluções produzidas pelo ED que consomem dados do próprio ecossistema e a quantidade total de soluções do ED	$X = N_i / N_t$ $N_i = \text{Número de soluções do ED que consomem dados do ecossistema}$ $N_t = \text{Número total de soluções do ecossistema}$	$0 \leq X \leq 1$ Quanto mais próximo do número 1 melhor	Dec - Decimal	Recursos	Autor (2019)
Indicador: Produtividade - Característica: Atividade								
P7	Produtividade	Número de eventos	Número de eventos que promovem os recursos do ED	$X = \sum_1^T N_T$ $N_T = \text{Número de eventos}$ $T = \text{Período de tempo}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o número de eventos realizados melhor	Un - Unidade	Recursos	Jansen (2014) Franco Bedoya et al. (2016)
P8	Produtividade	Número de materiais de referência	Número de materiais de referência para instruir os atores a usarem os recursos do ED	$X = N$ $N = \text{Número de materiais de referência}$	$X \geq 0$ Quanto maior o número de materiais melhor	-	Recursos	Autor (2019)

Fonte: A autora (2019)

4.2.4.2 Métricas de Robustez

As métricas de robustez são agrupadas por características. Dentre as características de robustez tem-se: Consistência Financeira, Garantia de Disponibilidade e Gestão de Riscos. Essas três características agrupam métricas que calculam dois aspectos diferentes que influenciam na saúde: o estado atual da robustez e a capacidade de se manter robusto ao longo do tempo.

O estado atual da robustez é calculado por métricas que avaliam se os dados possuem meios que garantam a disponibilidade contínua dos recursos em situações de instabilidade, a consistência financeira por meio do mapeamento de investidores (*i.e.* pontuais e permanentes), recursos financeiros produzidos, recebidos e gastos, como também pela variedade de formas de monetização (*e.g.* vendas de produtos digitais, *pay per use*, venda de informações de usuários).

A capacidade de se manter robusto é influenciada pela presença ou ausência da gestão de riscos no ecossistema. A existência de gestão no ecossistema revela a adoção de medidas preventivas caso haja perturbações no funcionamento do ecossistema. Já a presença de diferentes estruturas organizacionais (*e.g.* orientadas a modelos de negócio, centradas em *keystones*, centradas em plataforma) aumenta a robustez, pois eles protegem os atores contra perturbações externas. Somado a isso, contribuem para a gestão de riscos a presença de processos e sua documentação para guiar os atores na execução das atividades.

Tabela 5 – Métricas do Indicador de Robustez para Avaliar a Saúde de EDs

ID	Indicador	Métrica	Descrição	Medida e Fórmula	Interpretação	Unidade	Atributo	Referência Bibliográfica
Indicador: Robustez - Característica: Consistência Financeira								
R1	Robustez	Número de investidores privados permanentes	Número de atores privados que contribuem para o financiamento do ED de forma permanente	$X = \sum_1^T N_T$ $T = \text{Período de tempo}$ $N_T = \text{Número de investidores privados que financiam o ecossistema de forma permanente}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o número de atores melhor	Un - Unidade	Atores	Jansen (2014)
R2	Robustez	Número de investidores públicos permanentes	Número de atores públicos que contribuem para o financiamento do ED de forma permanente	$X = \sum_1^T N_T$ $T = \text{Período de tempo}$ $N_T = \text{Número de investidores públicos que financiam o ecossistema de forma permanente}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o número de atores melhor	Un - Unidade	Atores	Jansen (2014)
R3	Robustez	Número de investidores privados pontuais	Número de atores privados que contribuem para o financiamento do ED de forma pontual	$X = \sum_1^T N_T$ $T = \text{Período de tempo}$ $N_T = \text{Número de investidores privados que financiam o ecossistema de forma pontual}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o número de atores melhor	Un - Unidade	Atores	Jansen (2014)

R4	Robustez	Número de investidores públicos pontuais	Número de atores públicos que contribuem para o financiamento do ED de forma pontual	$X = \sum_1^T N_T$ $T = \text{Período de tempo}$ $N_T = \text{Número de investidores públicos que financiam o ecossistema de forma pontual}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o número de atores melhor	Un - Unidade	Atores	Jansen (2014)
R5	Robustez	Recursos financeiros investidos	Volume de recursos financeiros investido no ED	$X = \sum_1^T N_T$ $N_T = \text{Volume de recursos financeiros investido}$ $T = \text{Período de tempo}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o valor melhor	Un - Unidade	Atores	Jansen (2014)
R6	Robustez	Recursos financeiros produzidos	Volume de recursos financeiros produzido pelo ED	$X = \sum_1^T N_T$ $N_T = \text{Volume de recursos financeiros produzido}$ $T = \text{Período de tempo}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o valor melhor	Un - Unidade	Atores	Autor (2019)
R7	Robustez	Recursos financeiros gastos	Volume de recursos financeiros gasto pelo ED	$X = \sum_1^T N_T$ $N_T = \text{Volume de recursos financeiros gasto}$ $T = \text{Período de tempo}$	$X \leq P$ $P = V_{R6} + V_{R7}$ $V_{R6} = \text{Volume de recursos financeiros investido pelo ED}$ $V_{R7} = \text{Volume de recursos financeiros produzido pelo ED}$	Un - Unidade	Atores	Autor (2019)
R8	Robustez	Formas de monetização	Número de formas diferentes de monetização existentes no ED	$X = N$ $N = \text{Número de formas diferentes de monetização}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o número melhor	Un - Unidade	Recursos	Autor (2019)
Indicador: Robustez - Característica: Garantia de Disponibilidade								
R9	Robustez	Backup dos dados	Existência de backup dos dados no ED	$X = B$ $B = \text{Existência de backup dos dados}$	$X = \text{Sim} \mid \text{Não}$ Desejável que seja positivo	-	Recursos	Autor (2019)

R10	Robustez	Infraestrutura	Existência de infraestrutura que garanta a disponibilização dos recursos 24/7 no ED	$X = B$ B = Existência de infraestrutura que disponibilize os recursos 24/7	X = Sim Não Desejável que seja positivo	-	Recursos	Autor (2019)
R11	Robustez	Acesso por múltiplos mecanismos	Razão entre a quantidade de dados acessíveis por múltiplos mecanismos e a quantidade total de dados do ED.	$X = N_i / N_t$ N_i = Volume de dados que são acessíveis por mais de um mecanismo N_t = Volume total de dados	$0 \leq X \leq 1$ Quanto mais próximo do número 1 melhor	Dec - Decimal	Recursos	Autor (2019)
Indicador: Robustez - Característica: Gestão de Riscos								
R12	Robustez	Tipos de estruturas organizacionais	Número de tipos diferentes de estruturas organizacionais existentes no ED	$X = N$ N = Número de tipos diferentes de estruturas organizacionais	$X \geq P$ P = Parâmetro Quanto maior a diversidade melhor	Un - Unidade	Relações	Autor (2019)
R13	Robustez	Número de atores na gestão	Número de atores que atuam na gestão do ED	$X = \sum_1^T N_T$ N_T = Número de atores que atuam na gestão T = Período de tempo	$X \geq P$ P = Parâmetro Quanto maior o número de atores melhor	Un - Unidade	Atores	Autor (2019)
R14	Robustez	Existência de auditoria no ecossistema	Existência de auditoria no ED	$X = B$ B = Existência de auditoria	X = Sim Não Desejável que seja positivo	-	Atores	Autor (2019)
R15	Robustez	Existência de processos para execução de atividades	Existência de processos que auxiliam na execução das atividades dos atores	$X = B$ B = Existência de processos para execução das atividades do ED	X = Sim Não Desejável que seja positivo	-	Atores	Autor (2019)

R16	Robustez	Existência de documentação dos processos	Existência de documentação que descreva os processos das atividades exercidas pelos atores	X = B B = Existência de documentação	X = Sim Não Desejável que seja positivo	-	Atores	Autor (2019)
-----	----------	--	--	---	--	---	--------	--------------

Fonte: A autora (2019)

4.2.4.3 Métricas de Criação de Nicho

As métricas de criação de nicho são agrupadas pela característica Variedade. Essa característica agrupa métricas que calculam dois aspectos que influenciam na saúde: o estado atual da criação de nicho e a capacidade de promover novos nichos ao longo do tempo.

Esses dois aspectos são medidos juntos na característica de Variedade, pois podemos obter informações sobre a variedade de domínios representados pelos dados, como também os tipos de recursos produzidos, tipos de papéis exercidos pelos atores e os tipos de organizações públicas (*e.g. universidades, agências de fomento*) e privadas (*e.g. hospitais, multinacionais*) que fazem parte do ED. Essas informações fornecem indícios tanto do estado atual quanto da capacidade de criar novos nichos.

Tabela 6 – Métricas do Indicador de Criação de Nicho para Avaliar a Saúde de EDs

ID	Indicador	Métrica	Descrição	Medida e Fórmula	Interpretação	Unidade	Atributo	Referência Bibliográfica
Indicador: Criação de Nicho - Característica: Variedade								
C1	Criação de Nicho	Número de domínios representados pelos dados	Número de domínios representados pelos dados do ED	X = N N = Número de domínios representados pelos dados	X ≥ P P = Parâmetro Quanto maior o número de domínios melhor	Un - Unidade	Recursos	Dhungana et al. (2010)

C2	Criação de Nicho	Tipos de soluções	Número de tipos diferentes de soluções produzidas pelo ED	$X = N$ N = Número de tipos de soluções	$X \geq P$ P = Parâmetro Quanto maior a variedade de soluções melhor	Un - Unidade	Recursos	Jansen (2014) Franco Bedoya et al. (2016)
C3	Criação de Nicho	Tipos de organização privadas	Número de tipos diferentes de organizações privadas que fazem parte do ED	$X = N$ N = Número de tipos de organizações privadas	$X \geq P$ P = Parâmetro Quanto maior a variedade de organizações privadas melhor	Un - Unidade	Atores	Autor (2019)
C4	Criação de Nicho	Tipos de organizações públicas	Número de tipos diferentes de organizações públicas que fazem parte do ED	$X = N$ N = Número de tipos de organizações públicas	$X \geq P$ P = Parâmetro Quanto maior a variedade de organizações públicas melhor	Un - Unidade	Atores	Autor (2019)
C5	Criação de Nicho	Tipos de papéis	Número de tipos diferentes de papéis existentes no ED	$X = N$ N = Número de tipos diferentes de papéis	$X \geq P$ P = Parâmetro Quanto maior a variedade de papéis melhor	Un - Unidade	Papéis	Dhungana et al. (2010) Franco Bedoya et al. (2016)

Fonte: A autora (2019)

4.2.4.4 Métricas de Sustentabilidade

As métricas de sustentabilidade são agrupadas por características. Dentre as características de sustentabilidade tem-se: Habilidade de Regeneração, Equilíbrio de Esforços, Heterogeneidade, Engajamento e Qualidade. Essas cinco características agrupam métricas que calculam o estado atual da sustentabilidade do ED e a sua capacidade de se manter sustentável ao longo do tempo.

O estado atual da sustentabilidade é avaliado pela habilidade de regeneração do ED por meio da quantidade de atores ativos que estão no ecossistema desde o início e pela diferença entre o número de atores que entrou e saiu do ED em um período de tempo específico.

A capacidade de se manter sustentável pode ser avaliada com a medição do equilíbrio de esforços, pois a probabilidade do ecossistema se tornar instável se torna alta se um ator ou um pequeno grupo de atores que exercem uma atividade específica, saírem do ED. A heterogeneidade é outra característica importante pois a distribuição geográfica das parcerias revela a capacidade de sobrevivência do ecossistema caso outros locais não possam mais dar suporte. As parcerias são parte estratégica do ecossistema, pois são uma das formas de motivar os atores a contribuir (OLIVEIRA et al., 2018).

As métricas de engajamento dos participantes no ecossistema revelam a participação e as interações existentes em mídias sociais, as quais possuem um largo alcance nos dias de hoje, pois têm o poder de atrair e agregar mais atores para o ED. Além das citações em mídias sociais, o engajamento pode ser medido pela existência de canais de comunicação entre os atores, eventos de promoção do ED em outros ecossistemas e citações na literatura, como teses, artigos, patentes, entre outros (BOTELHO; OLIVEIRA, 2015). Utilizando essas métricas pode-se obter informações que revelem o grau de interesse do público e o potencial de crescimento do ED.

Por fim, as métricas de qualidade avaliam os dados produzidos pelo ecossistema em relação a aspectos que influenciam no consumo dos dados. São avaliados aspectos de segurança, acessibilidade e consistência, por exemplo. Com esses resultados conseguimos fornecer um panorama da qualidade dos dados consumidos pelos usuários e presentes nas soluções produzidas pelo ED.

Tabela 7 – Métricas do Indicador de Sustentabilidade para Avaliar a Saúde de EDs

ID	Indicador	Métrica	Descrição	Medida e Fórmula	Interpretação	Unidade	Atributo	Referência Bibliográfica
Indicador: Sustentabilidade - Característica: Habilidade de Regeneração								
S1	Sustentabilidade	Atores ativos desde o início do ED	Razão entre o número de atores ativos que estão desde o início do ecossistema e o total de atores ativos no ED	$X = N_i / N_t$ $N_i = \text{Número de atores ativos que participam do ecossistema desde o início}$ $N_t = \text{Número total de atores ativos que fazem parte do ecossistema}$	$0 \leq X \leq 1$ Quanto mais próximo do número 1 melhor	Dec - Decimal	Atores	Franco Bedoya et al. (2016)

S2	Sustentabilidade	Entrada de novos atores no ED	Número de novos atores que entraram no ecossistema	$X = \sum_1^T N_T$ $N_T = \text{Número de novos atores que entraram no ED}$ $T = \text{Período de tempo}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o número de novos atores melhor	Un - Unidade	Atores	Franco Bedoya et al. (2016)
S3	Sustentabilidade	Saída de atores do ED	Número de atores que interromperam sua participação no ED	$X = \sum_1^T N_T$ $N_T = \text{Número de atores que saíram do ecossistema}$ $T = \text{Período de tempo}$	$X \leq P$ $P = \text{Parâmetro}$ Quanto menor o número melhor	Un - Unidade	Atores	Autor (2019)
Indicador: Sustentabilidade - Característica: Equilíbrio de Esforços								
S4	Sustentabilidade	Centralização de esforços	Razão entre o número de atividades que somente um ator ou um pequeno grupo desempenha e o total de atividades do ED	$X = N_i / N_t$ $N_i = \text{Número de atividades exercidas por um ator ou um pequeno grupo}$ $N_t = \text{Número total de atividades do ED}$	$0 \leq X \leq 1$ Quanto mais próximo do número 0 melhor	Dec - Decimal	Atores	Franco Bedoya et al. (2016)
S5	Sustentabilidade	Distribuição de papéis	Razão entre o número de atores que exercem um papel específico e o total de atores do ED	$X = N_i / N_t$ $N_i = \text{Número de atores que exercem um papel específico}$ $N_t = \text{Número total de atores no ED}$	$0 \leq X \leq 1$ Quanto mais próximo do número 1 melhor	Dec - Decimal	Atores	Franco Bedoya et al. (2016)
Indicador: Sustentabilidade - Característica: Heterogeneidade								
S6	Sustentabilidade	Descentralização de parcerias	Razão entre o número de parcerias em mais de um país/estado/cidade e o total de parcerias do ED	$X = N_i / N_t$ $N_i = \text{Número de parcerias em mais de um país/estado/cidade}$ $N_t = \text{Número total de parcerias do ED}$	$0 \leq X \leq 1$ Quanto mais próximo do número 1 melhor	Dec - Decimal	Relacionamentos	Franco Bedoya et al. (2016)
Indicador: Sustentabilidade - Característica: Engajamento								

S7	Sustentabilidade	Citações sobre o ED em mídias sociais	Número de citações sobre o ED em mídias sociais	$X = \sum_1^T N_T$ $N_T =$ Número de citações publicadas $T =$ Período de tempo	$X \geq P$ $P =$ Parâmetro Quanto maior o número de citações melhor	Un - Unidade	Recursos	Franco Bedoya et al. (2016)
S8	Sustentabilidade	Canal de comunicação	Existência de canal de comunicação entre o consumidor e o produtor	$X = B$ $B =$ Existência de canal de comunicação entre os atores	$X =$ Sim Não Desejável que seja positivo	-	Atores	Autor (2019)
S9	Sustentabilidade	Eventos em ecossistemas externos	Participação dos atores do ecossistema em eventos de outros ecossistemas de dados	$X = B$ $B =$ Participação dos atores em eventos de outros ecossistemas	$X =$ Sim Não Desejável que seja positivo	-	Atores	Franco Bedoya et al. (2016)
S10	Sustentabilidade	Citações na Literatura	Número de trabalhos na literatura que citam o ED	$X = \sum_1^T N_T$ $N_T =$ Número de trabalhos na literatura $T =$ Período de tempo	$X \geq P$ $P =$ Parâmetro Quanto maior o número de trabalhos melhor	Un - Unidade	Recursos	Franco Bedoya et al. (2016)
Indicador: Sustentabilidade - Característica: Qualidade								
S11	Sustentabilidade	Facilidade de acesso	Os dados são facilmente acessados por download, API ou outro meio de acesso	$X = B$ $B =$ Se os dados são de fácil acesso	$X =$ Sim Não Desejável que seja positivo	-	Recursos	Lee et al. (2002)
S12	Sustentabilidade	Formato dos dados	Quantidade de formatos diferentes nos quais os dados são publicados	$X = N$ $N =$ Número de formatos diferentes	$X \geq P$ $P =$ Parâmetro Quanto maior o número de formatos melhor	Un - Unidade	Recursos	Zaveri et al. (2016)

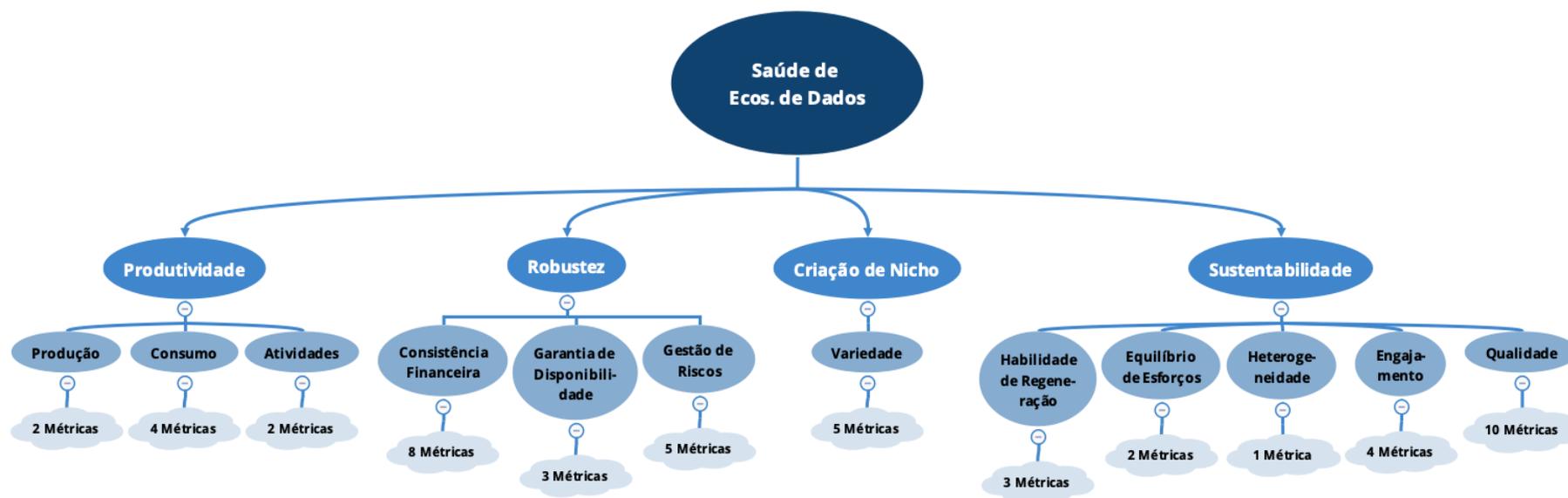
S13	Sustentabilidade	Remoção de dados	Há remoção de dados que não são acessados em um período de tempo	$X = B_T$ $B_T =$ Se os dados não acessados em um período de tempo T são removidos	X = Sim Não Desejável que seja positivo	-	Recursos	Autor (2019)
S14	Sustentabilidade	Uso de licença	Os dados são restringidos por alguma licença	$X = B$ $B =$ Se os dados não restringidos por alguma licença	X = Sim Não Desejável que seja positivo	-	Recursos	Zaveri et al. (2016)
S15	Sustentabilidade	Proteção contra acesso não autorizado	Dados sigilosos e/ou pessoais são protegidos contra acesso não autorizado	$X = B$ $B =$ Se dados sigilosos/pessoais são protegidos	X = Sim Não Desejável que seja positivo	-	Recursos	Zaveri et al. (2016)
S16	Sustentabilidade	Metadados	Os conjuntos de dados possuem metadados	$X = B$ $B =$ Se os conjuntos de dados possuem metadados	X = Sim Não Desejável que seja positivo	-	Recursos	Zaveri et al. (2016)
S17	Sustentabilidade	Refinamento dos dados	Há limpeza ou enriquecimento dos dados	$X = B$ $B =$ Se há remoção de dados inconsistentes	X = Sim Não Desejável que seja positivo	-	Recursos	Autor (2019)
S18	Sustentabilidade	Padronização dos dados	Há padronização na forma como os dados são apresentados	$X = B$ $B =$ Se há padronização na forma de apresentação dos dados	X = Sim Não Desejável que seja positivo	-	Recursos	Lee et al. (2002)
S19	Sustentabilidade	Confiabilidade dos dados	Existência de mecanismos para checar a veracidade dos dados	$X = B$ $B =$ Se há mecanismos para checar veracidade dos dados	X = Sim Não Desejável que seja positivo	-	Recursos	Lee et al. (2002)

S20	Sustentabilidade	Atualização dos dados	Razão entre a quantidade de dados atualizados e a quantidade total de dados do ED	$X = N_a / N_t$ $N_a =$ Quantidade de dados atualizados $N_t =$ Quantidade total de dados	$0 \leq X \leq 1$ Quanto mais próximo do número 1 melhor	Dec - Decimal	Recursos	Lee et al. (2002) Zaveri et al. (2016)
-----	------------------	-----------------------	---	---	---	---------------	----------	---

Fonte: A autora (2019)

Por fim, a Figura 11 detalha a estrutura e os elementos que compõem o FASED.

Figura 11 – FASED - Hierarquia e Elementos



Fonte: A autora (2019)

4.3 SÍNTESE DO CAPÍTULO

Este Capítulo apresentou a proposta de um *framework*, chamado FASED, para avaliação da saúde de Ecossistemas de Dados. Este *framework* foi construído a partir do estudo do estado da arte em Ecossistemas de Dados e pelo estudo *ad-hoc* de *frameworks* que avaliam a saúde de Ecossistemas de *Software* e de Negócios. O *framework* proposto adere aos padrões de construção de *framework* e modelos de referência. Somado a isso, o desenvolvimento do FASED foi guiado por um conjunto de indicadores, características e métricas referenciadas na literatura. Esse conjunto de elementos foi adaptado, melhorado, estendido, refinado e verificado em diversos ciclos até a versão atual apresentada neste trabalho.

O *framework* proposto é estruturado em 4 indicadores, 12 características (distribuídos entre os indicadores), 4 atributos e 49 métricas (distribuídas entre as características e de acordo com o atributo relacionado). O *framework* FASED forma um conjunto de elementos de caráter genérico e adaptável, permitindo que um grupo de atores consiga avaliar a saúde de EDs de acordo com o seu domínio.

No próximo Capítulo será apresentado um guia de instanciação do FASED, duas aplicações do *framework* em um cenário fictício e a análise do estado da saúde em períodos distintos.

5 GUIA DE INSTANCIACÃO DO FASED

5.1 COMO REALIZAR UMA AVALIAÇÃO

O *Framework* para Avaliação da Saúde de EDs fornece uma estrutura de avaliação da saúde por meio de indicadores, características e métricas. Como parte da realização de uma avaliação da saúde, devem ser definidas metas e o estado atual do ecossistema deve ser mapeado. Isso irá produzir uma avaliação completa da saúde e dará suporte ao acompanhamento do estado da saúde em diferentes períodos de tempo.

Alguns elementos do FASED serão mais aplicáveis a certos ecossistemas. Por exemplo, um ecossistema de dados centrado em uma plataforma de software e que lida com dados sensíveis, irá focar em métricas que avaliem aspectos sobre armazenamento, processamento e controle de acesso. Em contraponto, um ecossistema que apenas lida com dados públicos e não-sensíveis, poderá utilizar métricas focadas em criar oportunidades de consumo e na sua rentabilização.

Alguns ecossistemas produzem e administram dados como parte de sua tarefa principal, enquanto que outros geram dados como efeito colateral de suas atividades. Essas diferenças irão afetar na escolha dos indicadores, características e métricas, assim como na exclusão de métricas que não se aplicam à realidade do ED, como na derivação de novas métricas que mapeiem aspectos não abordados na versão genérica do FASED.

Os gestores dos EDs devem determinar as metas que eles querem alcançar para cada métrica. Conforme o ecossistema ganha valor a partir das ações tomadas para mitigar os riscos mapeados na avaliação da saúde, elas podem justificar investimentos adicionais, por exemplo. A avaliação da saúde e melhorias devem compor um processo iterativo.

É importante ressaltar que os resultados obtidos no final da avaliação não se destinam a eliminar papéis ou atividades. Por exemplo, um número baixo de domínios representados pelos dados não deve limitar o planejamento e publicação de novas soluções, e na criação de novas parcerias. Os resultados obtidos a partir da comparação entre o estado atual e as metas estabelecidas fornecem uma lista de lacunas que os atores da gestão do ED devem prestar atenção e focar em melhorias de acordo com as prioridades. É por meio desse mapeamento de pontos positivos e negativos que o ED consegue definir ações pontuais e mais eficazes para melhorar a saúde do ecossistema.

5.1.1 Processo de Avaliação

O processo a seguir descreve a abordagem recomendada para a realização de uma avaliação da saúde de EDs:

1. Identificar o(s) gestor(es) - a avaliação da saúde irá exigir provavelmente a contribuição de atores de diferentes tipos de papéis, entretanto essa avaliação deve ser

coordenada pelos gestores do ED.

2. Identificar atores-chave - quais atores do ecossistema precisam estar envolvidos para auxiliar na definição das metas e nas respostas de métricas específicas?
3. Identificar o domínio - o *framework* pode ser usado para avaliar qualquer domínio de ecossistemas de dados, por exemplo dados abertos, científicos, privados, de redes sociais, entre outros.
4. Definir período de tempo - a avaliação deve ser feita focada em um período de tempo que se queira avaliar a saúde, como por exemplo a saúde dos últimos 6 meses.
5. Definir métricas - após a identificação do domínio se faz necessária a definição de métricas que reflitam a realidade do ED. Para auxiliar nessa tarefa as métricas do FASED são genéricas e podem ser adaptadas para satisfazer as especificidades do ecossistema em avaliação.
6. Definir metas - após a definição das métricas, o(s) líder(es) e atores-chave devem participar da definição das metas para cada métrica, as quais servirão de base de comparação para o cálculo dos indicadores, e conseqüentemente, do estado da saúde.
7. Avaliar cada métrica - é necessário consultar os gestores e atores-chave para a coleta de cada métrica pois esses atores têm conhecimento sobre as diferentes características que serão avaliadas.
8. Desenvolver um plano de ação - de posse dos resultados e das metas, os gestores podem formular um plano para implantar melhorias.
9. Definir a data da próxima avaliação - a liderança e os gestores devem definir um período de tempo para que a próxima avaliação seja feita, e assim um panorama do estado da saúde seja feito para monitorar o progresso.

A ordem em que as ações são apresentadas reflete uma progressão aproximada e não significa que uma ação tenha maior valor ou prioridade em relação a outra. Além disso, os indicadores, características e métricas irão se relacionar entre si. Por exemplo, a realização de eventos para divulgar recursos produzidos pelo ecossistema, atrai novos atores para consumir os recursos do ED, como também contribuir na publicação de novas soluções baseadas nos dados. Somado a isso, aumenta as chances de criar novas parcerias e atrair novos investidores públicos ou privados. Da mesma forma, mapear a variedade de nichos existentes pode auxiliar criação de novas oportunidades, como eventos de divulgação do ecossistema direcionados a um público específico, criação de novos relacionamentos e desenvolvimento de soluções.

Um ED pode redefinir suas próprias prioridades baseado nos resultados da avaliação utilizando o FASED:

- Ao realizar uma avaliação, pode ser útil priorizar a verificação de certas características que são mais críticas para o ecossistema ou que são negligenciadas e os gestores querem mapear o estado atual; por exemplo, verificar como está a gestão de riscos, a qualidade dos dados, entre outros.
- Ao definir as metas e ter acesso aos resultados após a avaliação, os gestores podem desejar implantar mudanças tanto seguindo abordagem *top-down* (partindo da supervisão e estratégia) como também *bottom-up*.

5.2 CENÁRIO MOTIVACIONAL E INSTANCIAÇÃO DO FASED

Nesta Seção vamos apresentar um cenário motivacional e a instanciação do FASED para esse cenário em dois períodos diferentes. Um exemplo de Ecossistema de Dados é um Ecossistema de Dados Abertos Governamentais (EDGov), o qual é uma iniciativa de abertura de dados promovida por uma Administração Governamental fictícia da esfera nacional visando promover transparência e melhoria dos seus serviços.

Neste cenário, os dados seguem os oito princípios dos dados abertos governamentais¹. Os dados abertos governamentais (DAGov) são atualizados, coletados diretamente da fonte e disponibilizados para os usuários. Eles são completos, não proprietários e estruturados de maneira que possam ser processados de forma automática, *i.e. machine processable*. E somado a isso, os DAGov podem ser disponibilizados com restrições razoáveis de privacidade, segurança e privilégios.

Esse ED é centrado no Portal de Dados Abertos utilizado para publicar conjuntos de dados e materiais relacionados, *e.g.* documentação, metadados, *link* para aplicações. Atualmente, ele possui três conjuntos de dados, os quais são: (i) Censo Escolar, (ii) Aposentadorias Concedidas e (iii) Bolsas Concedidas para Ensino Superior. Os conjuntos de dados possuem informações sobre data da última atualização, data de criação, mantenedor, fonte, formatos disponíveis para acesso e download, licença, granularidade temporal (*e.g.* mensal, anual) e esfera do órgão responsável (*e.g.* nacional). Todos os conjuntos de dados são publicados sob a licença *Open Database License (ODbL)*², a qual permite o compartilhamento, adaptação e criação de novos recursos a partir desses dados, desde que citem a fonte original. A seguir, seguem algumas informações sobre o conteúdo de cada conjunto de dado:

- **Censo Escolar:** Esse conjunto de dados é atualizado anualmente no portal desde 1995 e é disponibilizado nos formatos CSV, JSON e por meio de uma API REST. Esse conjunto de dados traz informações sobre a disponibilidade de recursos relacionados à infraestrutura nas escolas públicas nacionais, como bibliotecas, banheiros, laboratórios, internet, pátio e quadras.

¹ <https://opengovdata.org>

² <https://opendefinition.org/licenses/odc-odbl/>

- **Aposentadorias Concedidas:** Esse conjunto de dados é atualizado mensalmente no portal desde 2015, é disponibilizado nos formatos CSV, JSON, ZIP e possui uma página HTML disponível para visualização dos dados com filtros e um mapa. Ele fornece dados como nome do beneficiário, tipo da aposentadoria, valor, data de início, data do último benefício recebido, estado e cidade.
- **Bolsas Concedidas para Ensino Superior:** Esse conjunto de dados é atualizado mensalmente no portal desde 2013 e é disponibilizado nos formatos CSV e JSON. Ele traz informações sobre os recursos financeiros pagos à alunos do curso superior de ensino, fornecendo dados sobre nome do aluno, início da vigência da bolsa, valor da bolsa, país, tipo da moeda, código da instituição de ensino superior, nome do curso, data da última parcela concedida, situação da bolsa (ativa ou inativa), modalidade (graduação, mestrado, doutorado ou pós-doutorado).

O Portal foi desenvolvido utilizando o CKAN, o qual é um *software open source* que provê a criação de portais de dados. Esse portal é hospedado no OpenShift³, plataforma que provê serviços de hospedagem para *Web sites*. Os componentes do OpenShift são organizados em um esquema de microserviços e utilizam serviços de *cloud* para armazenar os dados de forma descentralizada. Esse serviço de hospedagem utilizado para hospedar o portal segue o modelo *Freemium*, que é um modelo de negócio de oferta e demanda onde um serviço é disponibilizado sem cobrança, porém funcionalidades adicionais são cobradas.

Além disso, este EDGov é composto por diversos atores, como organizações privadas, públicas e indivíduos, os quais exercem um ou mais papéis. As organizações privadas, que são a *startup* SchoolAccess e a empresa INSR, atuam no ecossistema como consumidores dos dados e parceiros do ecossistema, fomentando e investindo no desenvolvimento de novos recursos. Nos últimos seis meses essas organizações privadas fizeram investimentos pontuais para promover eventos que divulgassem as aplicações desenvolvidas pelo ecossistema. Esses atores exercem os papéis de consumidor de dados e intermediário, este último porque fomentam a produção de novos recursos por meio de investimentos. As organizações públicas exercem papéis de regulador, produtor, publicador, investidor e consumidor no EDGov, desde a gestão do ecossistema até atividades do ciclo de vida dos conjuntos de dados (SILVA, 2019). Essas organizações públicas são o ministério da educação e economia. Elas atuam como canal principal de fornecimento dos dados, mas as universidades, agências de fomento, institutos de pesquisa ou secretarias governamentais, atuam como fontes primárias, pois são essas instituições as fontes dos dados que são enviados para os ministérios.

As parcerias existentes nesse ecossistema são descentralizadas e aumentam a colaboração entre os atores, pois envolvem promoção de eventos e desenvolvimento de novas

³ <https://www.openshift.com>

soluções. Os parceiros são localizados em diferentes estados e ao todo existem 10 parcerias, dentre as quais 6 são localizadas em outras localidades geográficas.

Os indivíduos são os cidadãos comuns que consomem os recursos disponibilizados pelo EDGov, publicam novas soluções baseadas nos dados do ED e também exercem papéis que contribuem desde o planejamento até o arquivamento dos conjuntos de dados, passando pela criação e refinamento (SILVA, 2019). Os cidadãos procuram informações sobre gastos governamentais para analisar a situação financeira do país, a infraestrutura e quais ações estão sendo tomadas para melhorar as áreas da educação, saúde, segurança, dentre outros. Os cidadãos comuns, além de consumirem os dados por meio de *download* dos conjuntos de dados, soluções (*e.g. aplicativos*) e serviços, eles também participam indiretamente do planejamento de novos conjuntos de dados e do refinamento, pois fornecem *feedback* sobre as melhorias que podem ser feitas nos dados e no portal para otimizar a experiência do usuário. Além de servir para sugestão de melhorias, os canais de *feedback* também coletam elogios e reclamações.

Já os indivíduos que atuam diretamente no ciclo de vida, são funcionários da administração governamental e desenvolvedores de *software*, por exemplo. Os funcionários governamentais trabalham nos órgãos públicos diretamente com a coleta e publicação dos dados. Já os desenvolvedores utilizam os dados para gerar soluções ou serviços, como aplicações móveis, *Web*, visualizações e análises que são utilizadas por outros atores.

O EDGov deste cenário motivacional possui diversos recursos além dos conjuntos de dados. Há a plataforma *Web*, que é o Portal de Dados Abertos já mencionado, e duas aplicações móveis, uma que disponibiliza visualizações (*e.g. gráficos*) sobre cada conjunto de dados do próprio ecossistema, contendo filtros que melhoram a navegação e descoberta de novas informações; a segunda aplicação é voltada para a área da educação, possui um mapa com informações sobre a infraestrutura das escolas públicas se baseando no conjunto de dados de Censo Escolar. Os serviços fornecidos por esses 2 aplicativos também são disponíveis no portal. Há também uma aplicação *Web* chamada "Diferentes Perspectivas pelo Mundo" publicada pela *startup* "OpenAccountability", que utiliza os dados disponibilizados no Portal de Dados Abertos para fazer comparações com dados abertos de outros países. O EDGov possui uma estrutura organizacional baseada na plataforma, a qual fornece infraestrutura e serviços para a disponibilização e consumo dos conjuntos de dados e outros recursos, atenuando assim problemas de interoperabilidade e usabilidade.

Para avaliar a saúde do EDGov seguimos o guia apresentado na Seção 5.1. Inicialmente, para o levantamento dos dados, foram consultados alguns atores que participam da gestão, e também atores-chaves que publicam e consomem os dados desse ecossistema. De acordo com o domínio do ED, dados abertos governamentais, foram derivadas novas métricas baseadas nas métricas genéricas propostas pelo FASED, com o objetivo de abranger as especificidades do domínio em análise. E, para estabelecer as metas que servirão de base para comparar com os resultados coletados, consultamos os atores-chave e os que atuam na

gestão. De posse das metas que a gestão pretendia atingir no período avaliado, foi possível realizar duas avaliações da saúde do EDGov que serviram de base para acompanhar o estado da saúde desse ecossistema e como insumos para mostrar a aplicabilidade e a relevância do FASED.

Nas Seções 5.2.1 e 5.2.3 serão apresentadas duas avaliações da saúde do EDGov utilizando o FASED. Para a avaliação, as métricas do FASED foram adaptadas para satisfazer os aspectos específicos de Ecossistemas de Dados Abertos Governamentais, como a existência de conjuntos de dados, portal de dados, aplicativos, eventos de promoção desses recursos, entre outros. De posse das métricas para avaliar a saúde do EDGov, foram agendadas as entrevistas com os atores fictícios. Durante a realização das entrevistas foram coletados os resultados para cada métrica e as metas esperadas pelos gestores no período de tempo avaliado. Como foram executadas duas avaliações, o processo de entrevistas para coletar os resultados e definir as metas, foi executado duas vezes também.

Com os resultados da primeira e segunda avaliação, conseguimos avaliar e mapear o estado da saúde dois períodos de tempo distintos. Detalhamos as descobertas na Seção 5.2.5.

5.2.1 Aplicação do FASED no EDGOV - 1ª Avaliação

Esta avaliação é a primeira avaliação da saúde feita neste EDGov. Os gestores do EDGov decidiram obter o panorama da saúde considerando o período dos últimos 6 meses, pois foi o período que houve mudanças significativas no ED, como a disponibilização de novas aplicações, inserção de um canal de comunicação entre os produtores e consumidores e melhorias na publicação dos dados, como inclusão de metadados, mais mecanismos de acesso e nova infraestrutura para o portal.

O FASED foi adaptado para ser aplicado ao contexto do EDGov e algumas métricas foram derivadas das originais para cobrir aspectos específicos do Ecossistemas de Dados Abertos Governamentais (*e.g.* a métrica P3 derivou as métricas P3-I e P3-II). Essa adaptação das métricas foi realizada após estudos das características do domínio a ser avaliado. Na Tabela 8 são apresentadas as métricas derivadas do FASED, os resultados de cada métrica, coletados por meio da entrevista com alguns atores do EDGov, e as metas que foram estabelecidas pelos gestores de acordo com o período de tempo analisado. Na Tabela 8, quando a meta estiver preenchida com o símbolo '-', significa que essa métrica não teve uma projeção estabelecida e o resultado coletado satisfaz.

Após a aplicação do *framework* e de posse dos resultados, é realizada a análise dos resultados na Seção 5.2.2. A análise dos resultados fornecerá indícios sobre quais características e aspectos estão impactando negativamente no ecossistema e quais precisam de mais atenção. De posse dos resultados, os gestores terão o panorama da saúde e insumos para tomar decisões, bem como os indícios sobre o estado de cada característica, como

produção e consumo de recursos, qualidade dos dados, equilíbrio de esforços, consistência financeira e outras características que influenciam a saúde.

Tabela 8 – Aplicação do FASED no EDGov - 1ª Avaliação

ID	Métrica	Resultado da 1ª Avaliação	Meta para o Período
Indicador: Produtividade - Característica: Produção			
P1	Quantos conjuntos de dados foram disponibilizados pelo ecossistema nos últimos 6 meses?	13 conjuntos de dados	13 conjuntos de dados
P2	Quantas soluções foram produzidas pelo ED nos últimos 6 meses?	2 aplicações	2 aplicações
Indicador: Produtividade - Característica: Consumo			
P3-I	Quantos acessos aos conjuntos de dados foram realizados nos últimos 6 meses?	6523 acessos	7000 acessos
P3-II	Quantos acessos ao portal foram realizados nos últimos 6 meses?	5102 acessos	5000 acessos
P4-I	Quantos downloads dos conjuntos de dados foram feitos nos últimos 6 meses?	3048 downloads	4000 downloads
P4-II	Quantos downloads dos aplicativos foram feitos nos últimos 6 meses?	700 downloads	500 downloads
P5	Quantas soluções de ecossistemas externos consumiram os dados do ED nos últimos 6 meses?	1 solução	5 soluções
P6	Qual a porcentagem de soluções produzidas pelo ED que consomem os dados do ecossistema?	100%	100%
Indicador: Produtividade - Característica: Atividade			
P7-I	Quantos eventos que utilizaram os conjuntos de dados de forma direta ou indireta foram feitos nos últimos 6 meses?	1	6 (1 por mês)
P7-II	Quantos eventos que utilizaram os aplicativos de forma direta ou indireta foram feitos nos últimos 6 meses?	2	6 (1 por mês)
P8-I	Há quantos materiais de referência que auxiliam no uso dos conjuntos de dados?	0	3 (1 por tipo de conj. de dados)
P8-II	Há quantos materiais de referência que auxiliam no uso dos aplicativos do ED?	1	2 (1 por aplicativo)
Indicador: Robustez - Característica: Consistência Financeira			
R1	Quantos atores de organizações privadas investiram no ED de forma permanente nos últimos 6 meses?	0	3
R2	Quantos atores de organizações públicas investiram no ED de forma permanente nos últimos 6 meses?	1	1
R3	Quantos atores de organizações privadas investiram no ED de forma pontual nos últimos 6 meses?	2	4
R4	Quantos atores de organizações públicas investiram no ED de forma pontual nos últimos 6 meses?	0	0
R5	Qual o volume de recursos financeiros investido no ED nos últimos 6 meses?	R\$ 500000 reais	R\$ 700000 reais

R6	Qual o volume de recursos financeiros produzido no ED nos últimos 6 meses?	R\$ 10000 reais	R\$ 20000 reais
R7	Qual o volume de recursos financeiros gasto no ED nos últimos 6 meses?	R\$ 480000 reais	R\$ 720000 reais
R8	Quantas formas de monetização existem no ED?	0	1
Indicador: Robustez - Característica: Garantia de Disponibilidade			
R9	Há backup dos dados?	Sim	Sim
R10	Há infraestrutura que garanta disponibilização dos dados 24/7?	Sim	Sim
R11	Qual a porcentagem de conjuntos de dados acessíveis por múltiplos mecanismos?	100%	100%
Indicador: Robustez - Característica: Gestão de Riscos			
R12	Quantas estruturas organizacionais estão presentes no ED?	1	1
R13	Quantos atores atuaram na gestão do ED nos últimos 6 meses?	4	4
R14	Há auditoria no ED?	Não	Sim
R15	Há processos bem definidos para execução das atividades no ED?	Não	Sim
R16	Há documentação desses processos?	Não	Sim
Indicador: Criação de Nicho - Característica: Variedade			
C1	Quantos domínios são representados pelos conjuntos de dados?	3	6
C2	Quantos tipos de soluções são produzidos no ED?	2	3
C3	Quantos tipos de organizações privadas fazem parte do ED?	2	2
C4	Quantos tipos de organizações públicas fazem parte do ED?	4	4
C5	Quantos tipos de papéis existem no ED?	5	5
Indicador: Sustentabilidade - Característica: Habilidade de Regeneração			
S1	Qual a porcentagem de atores ativos que estão no ED desde o início?	60%	90%
S2	Quantos atores foram adicionados no ED nos últimos 6 meses?	11 atores	20 atores
S3	Quantos atores saíram do ED nos últimos 6 meses?	7 atores	0 atores
Indicador: Sustentabilidade - Característica: Equilíbrio de Esforços			
S4	Qual a porcentagem de atividades que são exercidas por um ator ou um pequeno grupo?	15%	0%
S5-I	Qual a porcentagem de atores que exercem o papel de produtor de dados no ED?	20%	-
S5-II	Qual a porcentagem de atores que exercem o papel de publicador de dados no ED?	25%	-
S5-III	Qual a porcentagem de atores que exercem papéis de gestão no ED?	5%	-
S5-IV	Qual a porcentagem de atores que exercem papéis intermediários no ED?	50%	-
Indicador: Sustentabilidade - Característica: Heterogeneidade			

S6	Porcentagem de parcerias existentes em mais de um país/estado/cidade?	60%	-
Indicador: Sustentabilidade - Característica: Engajamento			
S7	Quantas citações sobre o ED foram publicadas em mídias sociais nos últimos 6 meses?	86	-
S8	Há algum canal de comunicação entre o consumidor e o produtor?	Sim	Sim
S9	Há participação de atores do ED em eventos de outros ecossistemas?	Sim	Sim
S10	Quantos trabalhos da literatura citam o ED e foram publicados nos últimos 6 meses?	2	-
Indicador: Sustentabilidade - Característica: Qualidade			
S11	Os conjuntos de dados são facilmente acessados?	Sim	Sim
S12	Em quantos formatos diferentes os conjuntos de dados são disponibilizados?	3	3
S13	Há remoção de conjunto de dados não acessados?	Não	Não
S14	Os dados são restringidos por alguma licença?	Sim	Sim
S15	Há dados sigilosos e/ou pessoais que são protegidos contra acesso não autorizado?	Não	Não
S16	Os conjuntos de dados possuem metadados?	Sim	Sim
S17	Há refinamento dos dados?	Sim	Sim
S18	Há padronização na forma como os dados são apresentados?	Sim	Sim
S19	Há mecanismos para checar a veracidade dos dados?	Não	Sim
S20	Qual a porcentagem de dados atualizados?	100%	100%

Fonte: A autora (2019)

5.2.2 Resultados da 1ª Avaliação da Saúde do EDGOV

A avaliação da saúde é iniciada após a coleta dos resultados e metas para cada métrica. A avaliação é dividida em 2 etapas, na primeira etapa é feita a comparação entre os resultados das métricas e as metas, e na segunda etapa são gerados os gráficos para auxiliar na visualização e interpretação dos dados. Na primeira etapa é feita a comparação entre os resultados coletados e as metas correspondentes definidas pelos gestores, a fim de checar o quanto as métricas estão cumprindo com o esperado. A saída dessa etapa é uma porcentagem para cada métrica que indica o nível de cumprimento do dado coletado em relação à meta estabelecida.

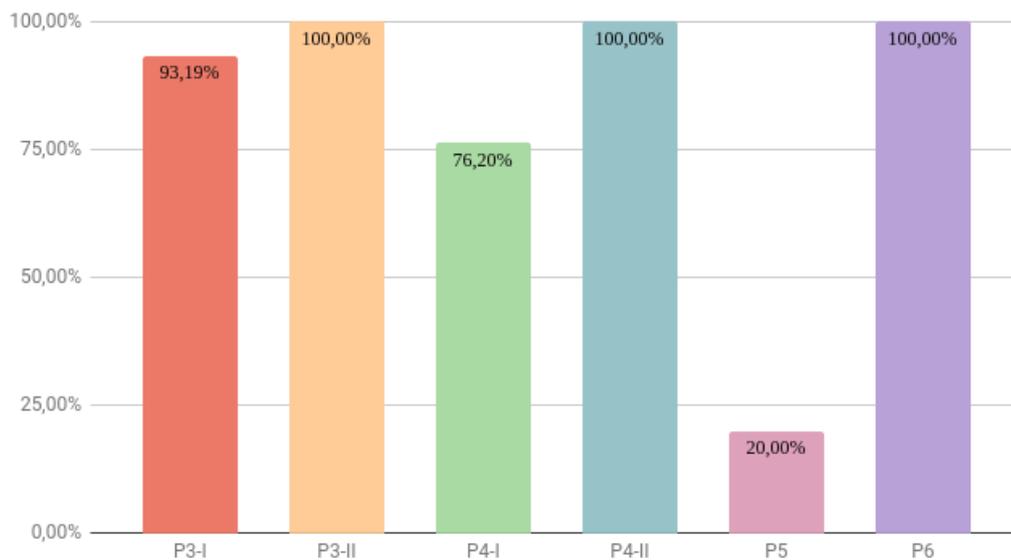
Na segunda etapa são feitas as análises dos resultados. A análise é feita em três passos, onde cada passo exibe o grau que o EDGov atende a cada nível de granularidade dos componentes analisados. Essa granularidade é dividida em características e métricas, indicadores de saúde e saúde geral.

A Figura 12 apresenta a comparação entre os resultados e as metas de cada métrica da característica Consumo. Cada barra do gráfico representa uma métrica dessa caracte-

rística e qual a porcentagem de cumprimento em relação à meta estabelecida. No exemplo ilustrado, as métricas P3-II, P4-II e P6 atingiram 100% das metas, enquanto que as métricas P3-I e P4-I atingiram 93.19% e 76.20%, respectivamente. Além disso, a métrica P5 atingiu somente 20% da meta, o que permite inferir que o aspecto analisado nessa métrica não está satisfazendo as expectativas dos gestores, e conseqüentemente necessita de mais atenção.

Figura 12 – Resultados da 1ª Avaliação - Métricas da Característica Consumo

Indicador: Produtividade - Característica: Consumo				
Métrica		Resultado	Meta	(%)
ID	Nome			
P3-I	Número de acessos aos conjuntos de dados	6523	7000	93,19%
P3-II	Número de acessos ao portal	5102	5000	100,00%
P4-I	Número de downloads de conjuntos de dados	3048	4000	76,20%
P4-II	Número de downloads dos aplicativos	700	500	100,00%
P5	Número de soluções de ecossistemas externos que consomem os dados do EDGov	1	5	20,00%
P6	Porcentagem de soluções produzidas pelo EDGov que consomem os próprios dados	100%	100%	100,00%



Fonte: A autora (2019)

A porcentagem de cada métrica facilita o cálculo das características, as quais são apresentadas a seguir. A porcentagem de cada característica é a média das porcentagens das métricas. A Figura 13 exibe a porcentagem de cumprimento de cada característica em relação ao indicador Produtividade, onde podemos ver que a característica Atividade foi a que menos atingiu os resultados esperados. Essa característica representa as atividades do ED que influenciam direta e indiretamente na produtividade, seja na realização de eventos que promovam os recursos, seja na existência de materiais que auxiliam no consumo das soluções do EDGov. Esse resultado gera indícios sobre a existência de lacunas e oportunidades nesses quesitos, pois a promoção de eventos internos e externos aumentam

a visibilidade e o potencial de agregar mais atores para produzir e consumir os dados, além de agregar valor aos produtos que estão sendo produzidos pelo EDGov.

Figura 13 – Resultados da 1ª Avaliação - Características do Indicador Produtividade



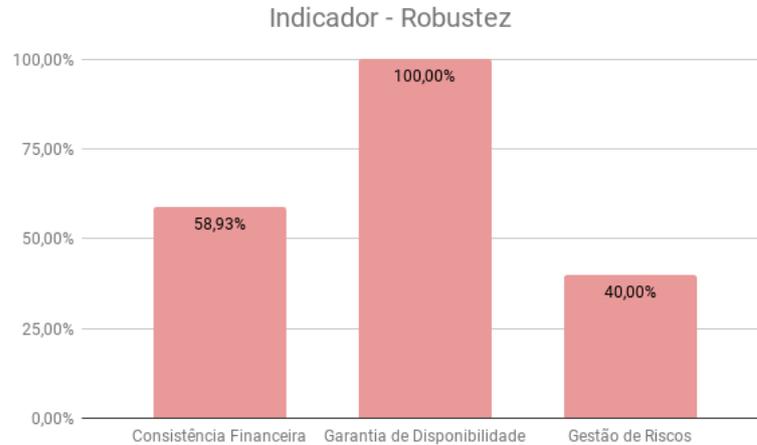
Fonte: A autora (2019)

Analisando os resultados do indicador Robustez na Figura 14, destacam-se as características de Consistência Financeira e Gestão de Riscos como as que menos atingiram o valor esperado, com 58.93% e 40%, respectivamente. A característica Consistência Financeira foi afetada pela ausência de investidores privados permanentes e pela ausência de formas de monetização. Esses são pontos que os gestores esperavam um resultado melhor mas não conseguiram atingir, e agora possuem indícios que está impactando negativamente na Robustez do ecossistema. Já a característica Gestão de Riscos indicou a ausência de ações que impactam na prevenção de riscos, como a ausência de auditoria, processos não estruturados e falta de documentação dos processos. Esses resultados fornecem indícios que existem lacunas na Gestão de Riscos e na Consistência Financeira, cabendo aos gestores após o recebimento do panorama, abordar os pontos deficientes e criar ações para melhorar a Robustez do ecossistema.

De acordo com a Figura 15 e com a Tabela 8, o indicador Criação de Nicho atingiu 83.33%, pois as métricas de Número de Domínios e Tipos de Soluções cumpriram somente 50% e 66.67%, respectivamente, da meta estabelecida.

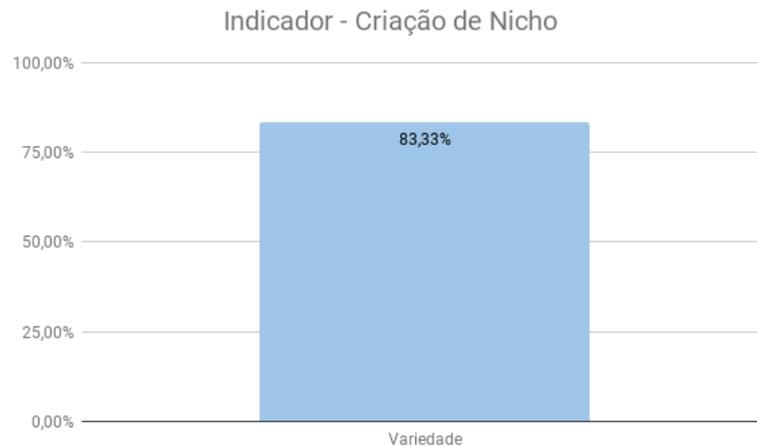
Por fim, a análise das características do indicador Sustentabilidade, Figura 16, mostra que a maioria das características atinge as metas ou estão próximas, como as características Habilidade de Regeneração e Qualidade, com 73.89% e 90%, respectivamente. Na característica Habilidade de Regeneração, as métricas que avaliam a porcentagem de atores ativos desde o início do ecossistema e a quantidade de atores adicionados ao ED atingiram mais de 50% da meta, mas ainda assim afetaram a visão geral da característica. Já na característica Qualidade, somente uma métrica não atingiu a meta, a qual detectou a ausência de mecanismos para checar a veracidade dos dados, entretanto os gestores co-

Figura 14 – Resultados da 1ª Avaliação - Características do Indicador Robustez



Fonte: A autora (2019)

Figura 15 – Resultados da 1ª Avaliação - Características do Indicador Criação de Nicho



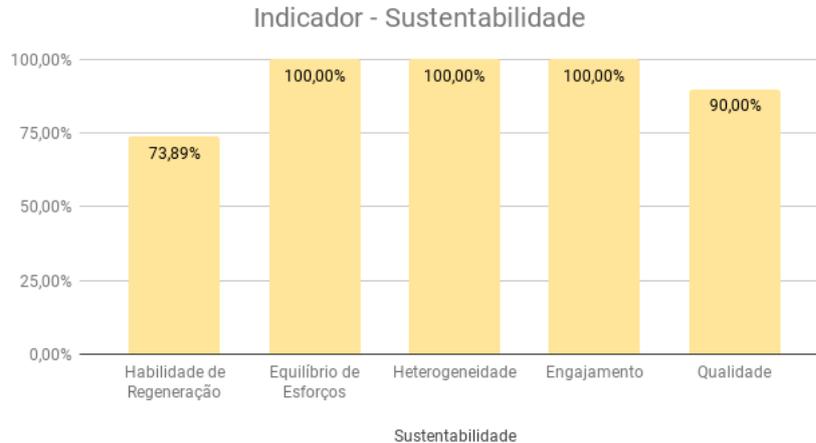
Fonte: A autora (2019)

locaram como meta ter esse tipo de mecanismo presente no EDGov. Com esses indícios os gestores possuem informações sobre quais pontos estão deficientes, e assim tomar ações específicas para melhorar a Sustentabilidade do EDGov.

Considerando o *framework* proposto neste trabalho, as métricas, características e indicadores possuem o mesmo peso na avaliação da saúde do ecossistema. A porcentagem de cada componente revela o quanto ele satisfaz as metas esperadas pelos gestores em um período de tempo, e de posse de todos os valores apresentados nessa análise, os gestores conseguem investigar quais aspectos merecem mais atenção de acordo com os objetivos do ecossistema naquele momento.

De acordo com a Figura 17, o panorama da saúde após a 1ª avaliação revela que os indicadores Produtividade, Robustez, Criação de Nicho e Sustentabilidade não cumpriram as metas estabelecidas pelos gestores, como já era esperado após as análises feitas

Figura 16 – Resultados da 1ª Avaliação - Características do Indicador Sustentabilidade



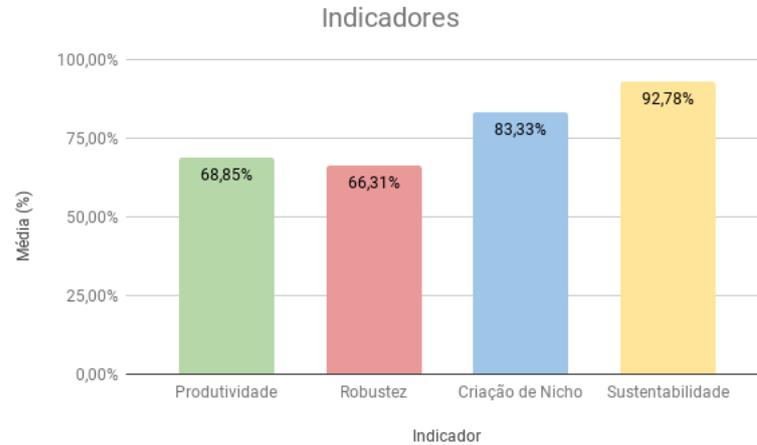
Fonte: A autora (2019)

anteriormente. O indicador Robustez foi o que menos atingiu os valores esperados, com 66.31%, e o indicador Sustentabilidade é o que gera indícios mais positivos sobre o seu estado, com 92.78%.

Para concluir, a Figura 18 traz a porcentagem geral da saúde com a média dos indicadores. Ao final desta 1ª avaliação, após a análise da estrutura do EDGov, por meio da medição de aspectos concretos relacionados ao consumo, produção, investimento e fomento dos recursos, atores e relacionamentos, fornecemos o panorama da saúde do EDGov como sendo 77.47% saudável. Com o panorama fornecido, a gestão conseguirá interferir em pontos críticos do ecossistema que estavam sendo negligenciados e assim tomar ações que mitiguem riscos e promovam a saúde do ecossistema.

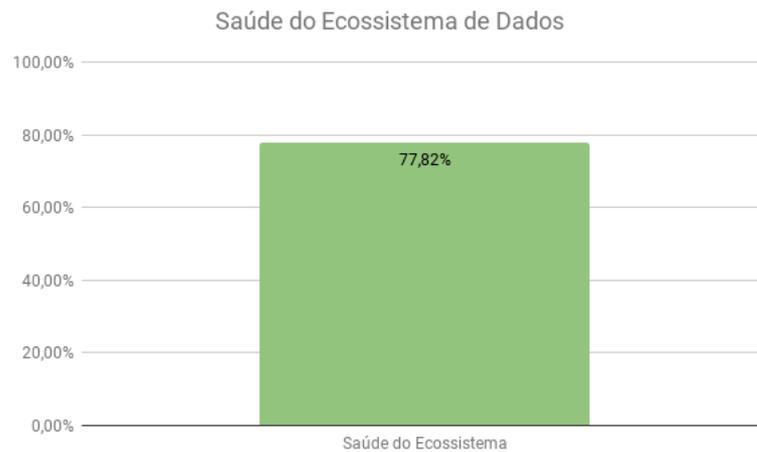
De posse desse valor, podemos fazer um acompanhamento e traçar um histórico do estado da saúde. Esta foi a primeira avaliação, e será feita uma segunda avaliação, detalhada nas Seções 5.2.3 e 5.2.4, para assim conseguirmos comparar o estado da saúde do ecossistema na Seção 5.2.5.

Figura 17 – Resultados da 1ª Avaliação - Visão Geral dos Indicadores



Fonte: A autora (2019)

Figura 18 – Resultados da 1ª Avaliação - Saúde Geral do EDGov



Fonte: A autora (2019)

5.2.3 Aplicação do FASED no EDGOV - 2ª Avaliação

De posse do panorama da saúde resultante da primeira avaliação, os gestores puderam ter uma visão de quais aspectos do ecossistema precisavam de mais atenção e assim tomar ações para melhorar a saúde do EDGov. A segunda avaliação da saúde do EDGov foi realizada 6 meses após a primeira avaliação. Na segunda avaliação utilizamos como parâmetro novas metas coletadas com os gestores, pois algumas variáveis do ecossistema mudaram, e conseqüentemente os objetivos e projeções. Foram aplicadas as mesmas métricas e os resultados podem ser encontrados na Tabela 9. O mesmo processo de avaliação executado na primeira avaliação também foi seguido aqui. Na Seção 5.2.4 serão apresentadas as análises dos resultados.

Tabela 9 – Aplicação do FASED no EDGov - 2ª Avaliação

ID	Métrica	Resultado da 2ª Avaliação	Meta para o Período
Indicador: Produtividade - Característica: Produção			
P1	Quantos conjuntos de dados foram disponibilizados pelo ecossistema nos últimos 6 meses?	18 conjuntos de dados	13 conjuntos de dados
P2	Quantas soluções foram produzidas pelo ED nos últimos 6 meses?	1 aplicação	2 aplicações
Indicador: Produtividade - Característica: Consumo			
P3-I	Quantos acessos aos conjuntos de dados foram realizados nos últimos 6 meses?	16204 acessos	7000 acessos
P3-II	Quantos acessos ao portal foram realizados nos últimos 6 meses?	12673 acessos	5000 acessos
P4-I	Quantos downloads dos conjuntos de dados foram feitos nos últimos 6 meses?	8127 downloads	4000 downloads
P4-II	Quantos downloads dos aplicativos foram feitos nos últimos 6 meses?	2741 downloads	500 downloads
P5	Quantas soluções de ecossistemas externos consumiram os dados do ED nos últimos 6 meses?	3 solução	5 soluções
P6	Qual a porcentagem de soluções produzidas pelo ED que consomem os dados do ecossistema?	100%	100%
Indicador: Produtividade - Característica: Atividade			
P7-I	Quantos eventos que utilizaram os conjuntos de dados de forma direta ou indireta foram feitos nos últimos 6 meses?	8	6 (1 por mês)
P7-II	Quantos eventos que utilizaram os aplicativos de forma direta ou indireta foram feitos nos últimos 6 meses?	8	6 (1 por mês)
P8-I	Há quantos materiais de referência que auxiliam no uso dos conjuntos de dados?	4	3 (1 por tipo de conj. de dados)
P8-II	Há quantos materiais de referência que auxiliam no uso dos aplicativos do ED?	3	2 (1 por aplicativo)
Indicador: Robustez - Característica: Consistência Financeira			
R1	Quantos atores de organizações privadas investiram no ED de forma permanente nos últimos 6 meses?	1	3
R2	Quantos atores de organizações públicas investiram no ED de forma permanente nos últimos 6 meses?	1	1
R3	Quantos atores de organizações privadas investiram no ED de forma pontual nos últimos 6 meses?	5	4
R4	Quantos atores de organizações públicas investiram no ED de forma pontual nos últimos 6 meses?	0	0
R5	Qual o volume de recursos financeiros investido no ED nos últimos 6 meses?	R\$ 1000000 reais	R\$ 700000 reais
R6	Qual o volume de recursos financeiros produzido no ED nos últimos 6 meses?	R\$ 20000 reais	R\$ 20000 reais
R7	Qual o volume de recursos financeiros gasto no ED nos últimos 6 meses?	R\$ 700000 reais	720000

R8	Quantas formas de monetização existem no ED?	1	1
Indicador: Robustez - Característica: Garantia de Disponibilidade			
R9	Há backup dos dados?	Sim	Sim
R10	Há infraestrutura que garanta disponibilização dos dados 24/7?	Sim	Sim
R11	Qual a porcentagem de conjuntos de dados acessíveis por múltiplos mecanismos?	100%	100%
Indicador: Robustez - Característica: Gestão de Riscos			
R12	Quantas estruturas organizacionais estão presentes no ED?	1	1
R13	Quantos atores atuaram na gestão do ED nos últimos 6 meses?	15	4
R14	Há auditoria no ED?	Não	Sim
R15	Há processos bem definidos para execução das atividades no ED?	Não	Sim
R16	Há documentação desses processos?	Não	Sim
Indicador: Criação de Nicho - Característica: Variedade			
C1	Quantos domínios são representados pelos conjuntos de dados?	4	6
C2	Quantos tipos de soluções são produzidos no ED?	3	3
C3	Quantos tipos de organizações privadas fazem parte do ED?	2	2
C4	Quantos tipos de organizações públicas fazem parte do ED?	4	4
C5	Quantos tipos de papéis existem no ED?	5	5
Indicador: Sustentabilidade - Característica: Habilidade de Regeneração			
S1	Qual a porcentagem de atores ativos que estão no ED desde o início?	60%	90%
S2	Quantos atores foram adicionados no ED nos últimos 6 meses?	30 atores	20 atores
S3	Quantos atores saíram do ED nos últimos 6 meses?	2 atores	0 atores
Indicador: Sustentabilidade - Característica: Equilíbrio de Esforços			
S4	Qual a porcentagem de atividades que são exercidas por um ator ou um pequeno grupo?	15%	0%
S5-I	Qual a porcentagem de atores que exercem o papel de produtor de dados no ED?	20%	-
S5-II	Qual a porcentagem de atores que exercem o papel de publicador de dados no ED?	20%	-
S5-III	Qual a porcentagem de atores que exercem papéis de gestão no ED?	10%	-
S5-IV	Qual a porcentagem de atores que exercem papéis intermediários no ED?	50%	-
Indicador: Sustentabilidade - Característica: Heterogeneidade			
S6	Porcentagem de parcerias existentes em mais de um país/estado/cidade?	75%	-
Indicador: Sustentabilidade - Característica: Engajamento			

S7	Quantas citações sobre o ED foram publicadas em mídias sociais nos últimos 6 meses?	487	-
S8	Há algum canal de comunicação entre o consumidor e o produtor?	Sim	Sim
S9	Há participação de atores do ED em eventos de outros ecossistemas?	Sim	Sim
S10	Quantos trabalhos da literatura citam o ED e foram publicados nos últimos 6 meses?	2	-
Indicador: Sustentabilidade - Característica: Qualidade			
S11	Os conjuntos de dados são facilmente acessados?	Sim	Sim
S12	Em quantos formatos diferentes os conjuntos de dados são disponibilizados?	3	3
S13	Há remoção de conjunto de dados não acessados?	Não	Não
S14	Os dados são restringidos por alguma licença?	Sim	Sim
S15	Há dados sigilosos e/ou pessoais que são protegidos contra acesso não autorizado?	Não	Não
S16	Os conjuntos de dados possuem metadados?	Sim	Sim
S17	Há refinamento dos dados?	Sim	Sim
S18	Há padronização na forma como os dados são apresentados?	Sim	Sim
S19	Há mecanismos para checar a veracidade dos dados?	Não	Sim
S20	Qual a porcentagem de dados atualizados?	100%	100%

Fonte: A autora (2019)

5.2.4 Resultados da 2ª Avaliação da Saúde do EDGOV

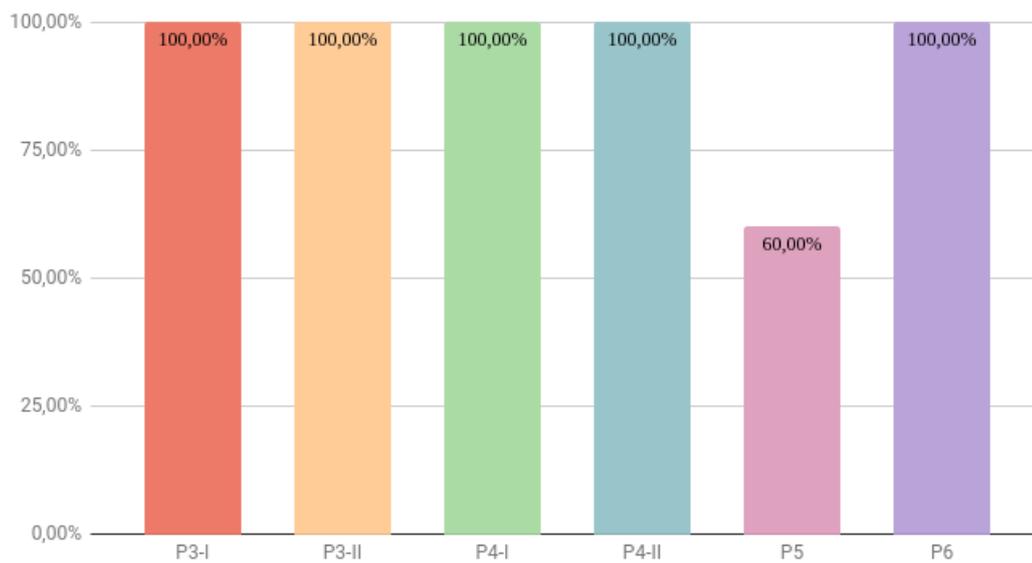
A segunda avaliação da saúde é iniciada após a coleta dos resultados e é feita a comparação dos resultados obtidos com os resultados da primeira avaliação. Dessa forma, torna-se possível traçar o panorama da saúde do EDGov após as 2 avaliações utilizando o FASED.

A Figura 19 apresenta a comparação entre os resultados e as metas de cada métrica da característica Consumo. Cada barra do gráfico representa uma métrica dessa característica e qual a porcentagem de cumprimento em relação à meta estabelecida. No exemplo ilustrado, as métricas P3-I, P3-II, P4-I, P4-II e P6 atingiram 100% das metas, enquanto que a métrica P5, sobre consumo dos dados do EDGov por ecossistemas externos, foi a que menos atingiu a meta, com 60%, o que permite inferir que os gestores esperam que os recursos do EDGov tenham mais visibilidade e sejam utilizados para agregar valor a soluções de outros ecossistemas.

A porcentagem de cada métrica possibilita o cálculo das características, que são apresentadas a seguir agrupadas por indicador. A porcentagem de cada característica é a média das porcentagens das métricas. A Figura 20 exhibe a porcentagem de atendimento de cada característica do indicador Produtividade, onde podemos ver que a característica Produção foi a que menos atingiu o esperado. Essa característica tem o objetivo de medir

Figura 19 – Resultados da 2ª Avaliação - Métricas da Característica Consumo

Indicador: Produtividade - Característica: Consumo				
Métrica		Resultado	Meta	(%)
ID	Nome			
P3-I	Número de acessos aos conjuntos de dados	16204	7000	100,00%
P3-II	Número de acessos ao portal	12673	5000	100,00%
P4-I	Número de downloads de conjuntos de dados	8127	4000	100,00%
P4-II	Número de downloads dos aplicativos	2741	500	100,00%
P5	Número de soluções de ecossistemas externos que consomem os dados do EDGov	3	5	60,00%
P6	Porcentagem de soluções produzidas pelo EDGov que consomem os próprios dados	100%	100%	100,00%



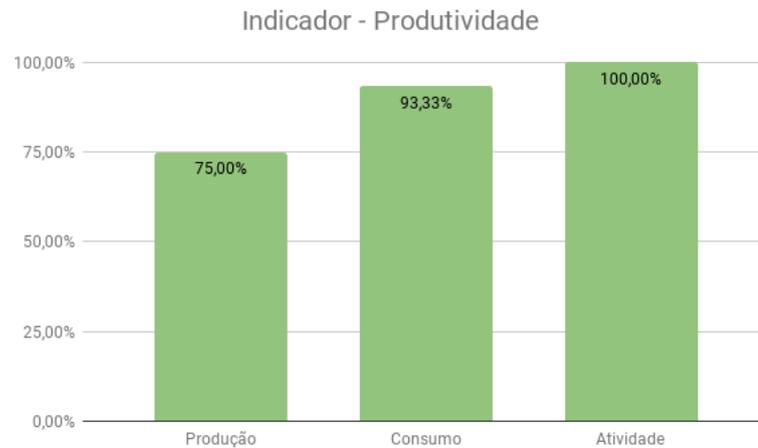
Fonte: A autora (2019)

o volume de recursos produzidos e foram coletados resultados de duas métricas, referentes a quantidade de conjuntos de dados e de soluções disponibilizados. A primeira, métrica P1, atingiu 100% da meta, mas a segunda métrica P2, sobre a quantidade de soluções produzidas nos últimos 6 meses, só atingiu 50% da meta. Com esse resultado, há indícios que existem lacunas na produção de soluções e que os gestores precisam prestar mais atenção nessa característica se quiserem otimizar a produção de recursos.

Analisando os resultados do indicador Robustez na Figura 21, destacam-se as características de Consistência Financeira e Gestão de Riscos as que menos atingiram o valor esperado, com 91.67% e 40%, respectivamente. A característica Consistência Financeira foi afetada pelo baixo número de investidores privados permanentes. Já os aspectos que afetaram a Gestão de Riscos na primeira avaliação, permanecem afetando na segunda avaliação, que são a ausência de ações que impactam na prevenção de riscos, como a ausência de auditoria, processos definidos e documentação dos processos.

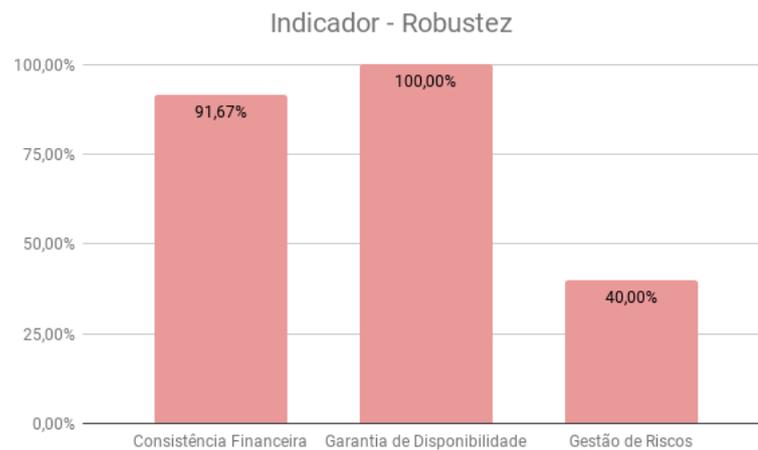
De acordo com a Figura 22 e com a Tabela 9, o indicador Criação de Nicho atingiu 93.33%, pois é esperado que sejam cobertos 6 domínios diferentes pelos conjuntos de

Figura 20 – Resultados da 2ª Avaliação - Características do Indicador Produtividade



Fonte: A autora (2019)

Figura 21 – Resultados da 2ª Avaliação - Características do Indicador Robustez



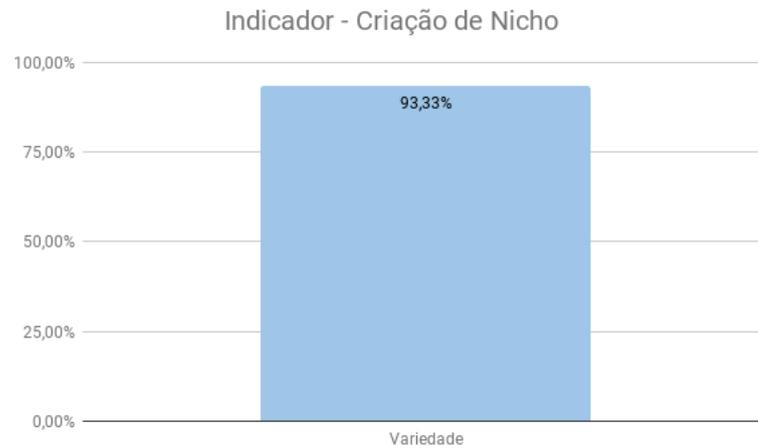
Fonte: A autora (2019)

dados, mas no período da avaliação só tinha 4.

Por fim, a análise das características do indicador Sustentabilidade, ver valores na Figura 23, mostram que a maioria atinge as metas ou está próxima, como as características Habilidade de Regeneração e Qualidade, com 88.89% e 90%, respectivamente. Na característica Habilidade de Regeneração, a métrica da porcentagem de atores ativos desde o início do ecossistema atingiu 66.67% da meta, mas ainda assim afeta a visão geral da característica. Já na característica Qualidade, somente uma métrica não atingiu a meta, que foi a métrica de verificar a existência de mecanismos para checar a veracidade dos dados, como já detectado na primeira avaliação.

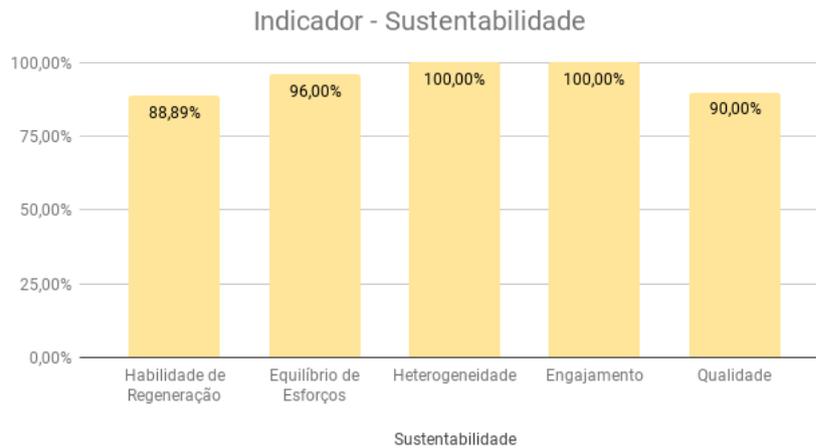
Considerando o FASED, os indicadores propostos possuem o mesmo peso na avaliação da saúde do ecossistema. A porcentagem de cada indicador revela o quanto ele satisfaz as metas esperadas pelos gestores em um período de tempo, e de posse de todos os valores

Figura 22 – Resultados da 2ª Avaliação - Características do Indicador Criação de Nicho



Fonte: A autora (2019)

Figura 23 – Resultados da 2ª Avaliação - Características do Indicador Sustentabilidade



Fonte: A autora (2019)

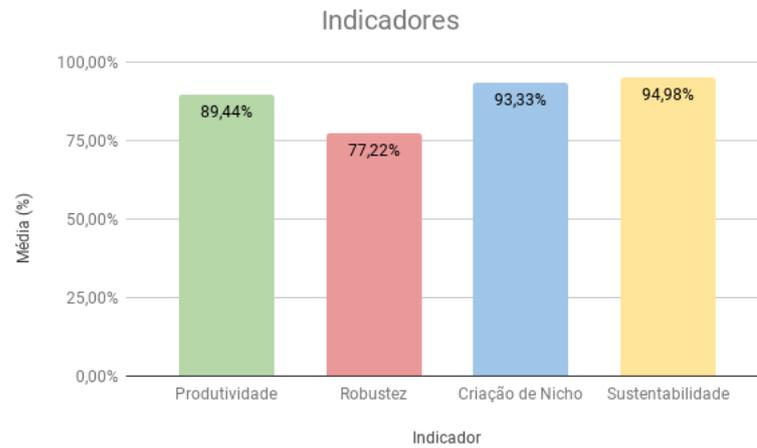
apresentados nessa análise, os gestores conseguem investigar quais aspectos merecem mais atenção de acordo com os objetivos do ecossistema naquele momento.

De acordo com a Figura 24, o panorama da saúde após a 2ª avaliação revela que os indicadores Produtividade, Robustez, Criação de Nicho e Sustentabilidade não cumpriram as metas estabelecidas pelos gestores. O panorama da saúde do EDGov foi definido através da análise da estrutura de sua rede de atores, recursos e outros componentes. O indicador Robustez foi o que menos atingiu, com 77.22%, e o indicador Sustentabilidade é o que mais está alinhado com as metas, com 94.98%, indicando uma melhora considerável no estado das suas características.

Para concluir, a Figura 25 traz a porcentagem geral da saúde com a média dos indicadores. De posse desse valor podemos fazer um acompanhamento e traçar um histórico da avaliação do estado da saúde do EDGov. Esta foi a segunda avaliação e será feita uma

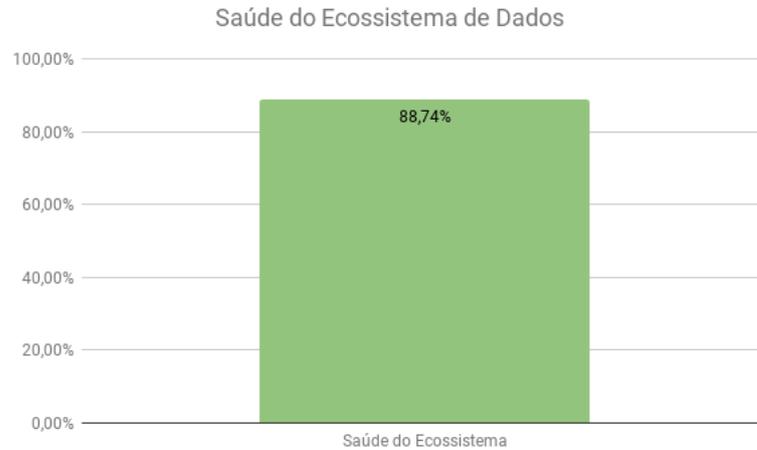
comparação entre as análises das duas avaliações para traçar o estado da saúde do EDGov em dois períodos de tempo distintos. Ver mais detalhes na Seção 5.2.5.

Figura 24 – Resultados da 2ª Avaliação - Visão Geral dos Indicadores



Fonte: A autora (2019)

Figura 25 – Resultados da 2ª Avaliação - Saúde Geral do EDGov



Fonte: A autora (2019)

5.2.5 Estado da Saúde em Períodos Distintos

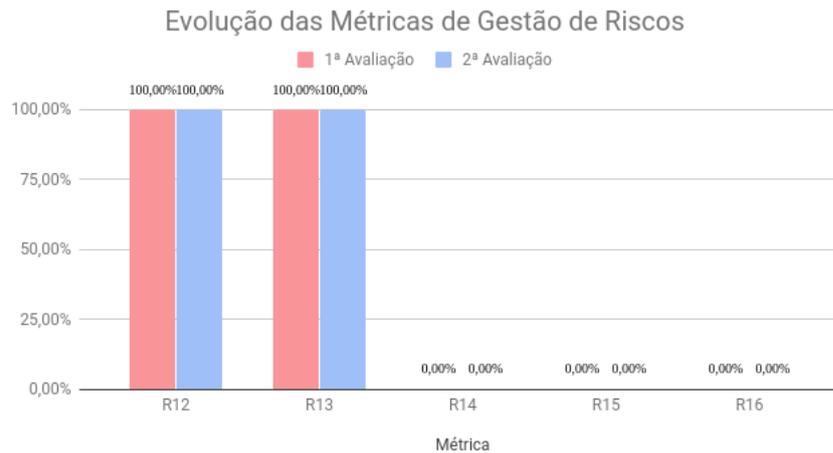
Com os resultados das duas avaliações é possível fazer um comparativo e analisar a saúde do EDGov em dois períodos de tempo distintos. Gerar o panorama da saúde, como foi feito nas Seções 5.2.2 e 5.2.4 é relevante para os gestores pois revelam aspectos que estão impactando negativamente na saúde do ED. Por isso, traçar o histórico da saúde por meio de mais de uma avaliação, mostra o estado da saúde e provê insumos para ratificar a importância de se avaliar a saúde e que ações foram tomadas para mitigar os riscos expostos nas análises pós avaliação.

Abaixo serão apresentados 4 gráficos para exemplificar, em diferentes granularidades, o estado da saúde, por meio das métricas de duas características, dos indicadores e da saúde de forma geral. Para ilustrar a evolução das métricas trouxemos dois casos distintos, o primeiro caso onde não houve alteração dos resultados da característica, entre as duas avaliações, e o segundo caso onde houve uma evolução significativa do estado da característica. Os casos são ilustrados nas Figuras 26 e 27, respectivamente.

A Figura 26 detalha a característica Gestão de Riscos com o seu conjunto de métricas. Analisando os resultados da primeira avaliação, detectamos que houve 3 métricas que não atingiram as metas, e mesmo assim, na segunda avaliação essa situação não mudou. Existem vários fatores que podem explicar porque na segunda avaliação elas ainda não atingiram os valores esperados, algumas justificativas podem ser: ações para alcançar os valores esperados estão em andamento, falta de mão de obra específica para executar essas atividades, entre outros. Esse gráfico representa um caso onde não houve alteração no estado da característica, mesmo havendo indícios que alguns aspectos estavam afetando negativamente a saúde do EDGov. Outras características do FASED que não sofreram alteração foram: Garantia de Disponibilidade, Heterogeneidade, Engajamento, todos permaneceram atingindo 100%, e a característica Qualidade permaneceu em 90%.

Figura 26 – Avaliação do Estado da Saúde - Métricas da Característica Gestão de Riscos

Indicador: Robustez - Característica: Gestão de Riscos			
Métrica		1ª Avaliação	2ª Avaliação
ID	Nome		
R12	Número de Estruturas Organizacionais	100,00%	100,00%
R13	Número de Atores na Gestão	100,00%	100,00%
R14	Existência de Auditoria	0,00%	0,00%
R15	Existência de Processos	0,00%	0,00%
R16	Documentação de Processos	0,00%	0,00%



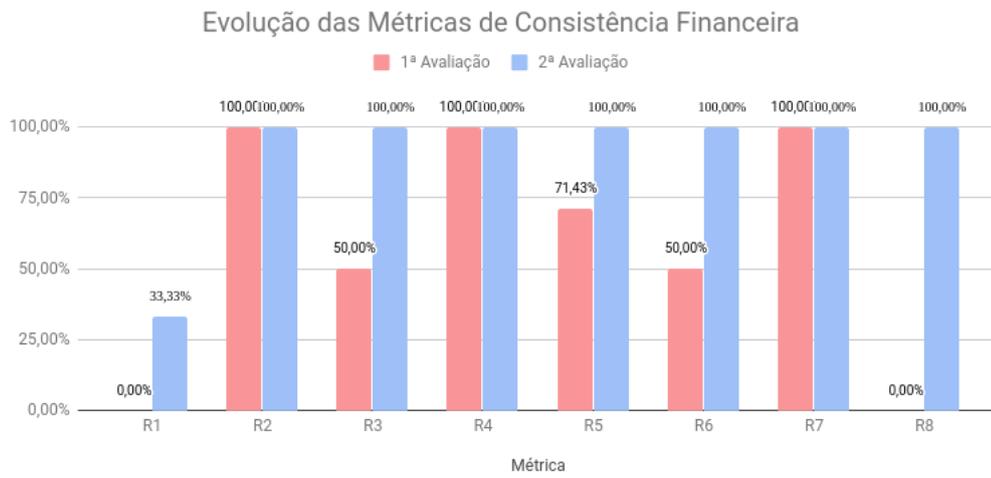
Fonte: A autora (2019)

Por outro lado, a Figura 27 retrata uma situação diferente, onde houve mudança nos resultados das métricas da característica Consistência Financeira. Analisando os valores das métricas R1, R3, R5, R6 e R8 nas duas avaliações, é possível notar a diferença nos resultados. As métricas R1 e R3, as quais são relacionadas com atração de investidores, revelam que após a primeira avaliação houve uma preocupação com esses aspectos e o EDGov conseguiu atrair mais investidores privados, tanto de forma permanente quanto pontual. Da mesma forma aconteceu com o indicador R5, o qual reflete o volume de recursos financeiros investidos, que conseguiu alcançar a meta esperada pelos gestores. Já a métrica R6 representa o volume de recursos financeiros produzidos pelo próprio EDGov, o qual teve um discreto aumento e passou de 50% para 100%. Por fim, a métrica R8 foi a métrica que mais evoluiu, antes era 0%, mas na segunda avaliação atingiu 100% da meta ao adotar uma forma de monetização. As outras métricas, R2, R4 e R7, permaneceram atingindo as metas estipuladas pelos gestores. No total, a característica Consistência Financeira passou de 58.93% para 91.67%.

Outras características que também exibiram evolução nas suas métricas foram: Consumo, passou de 81.56% para 93.33%; Variedade, passou de 83.33% para 93.33%; e Habilidade de Regeneração, passou de 73.89% para 88.89%. A característica que mais chamou atenção quanto a sua evolução foi a de Atividade, pois ela passou de 25% para 100%. Essa característica detecta como está a capacidade de promoção da produtividade no ED, tanto na forma de eventos para aumentar a visibilidade do ecossistema e atrair novos atores, como também na disponibilização de materiais que auxiliem no consumo e

Figura 27 – Avaliação do Estado da Saúde - Métricas da Característica Consistência Financeira

Indicador: Robustez - Característica: Consistência Financeira			
Métrica		1ª Avaliação	2ª Avaliação
ID	Nome		
R1	Número de Investidores Privados Permanente	0,00%	33,33%
R2	Número de Investidores Públicos Permanente	100%	100%
R3	Número de Investidores Privados Pontual	50%	100%
R4	Número de Investidores Públicos Pontual	100%	100%
R5	Recursos Financeiros Investidos	71,43%	100,0%
R6	Recursos Financeiros Produzidos	50%	100,00%
R7	Recursos Financeiros Gastos	100,00%	100,00%
R8	Número de Formas de Monetização	0%	100%



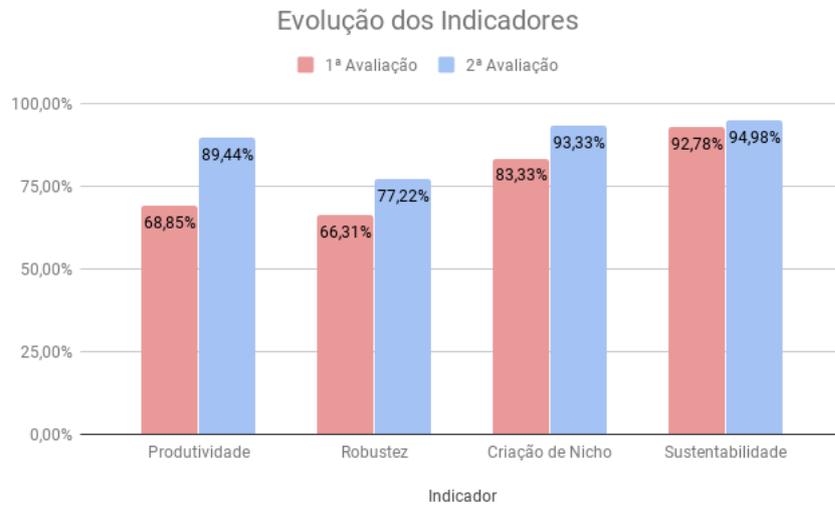
Fonte: A autora (2019)

produção de recursos. Analisando a evolução dessa característica e do aumento do número de investidores e de investimentos, podemos enxergar uma relação de causa e efeito, pois a exposição do EDGov por meio dos eventos e materiais mostram o seu potencial e, conseqüentemente, atraem mais atores para fomentar e investir no ecossistema.

A Figura 28 ilustra a evolução de cada indicador após as duas avaliações. É notável, como já falamos anteriormente, a evolução do indicador Produtividade ao endereçar esforços para a característica Atividade. Esse foi o indicador com maior evolução. O indicador Robustez evoluiu, mas ainda assim, após a segunda avaliação, está com a porcentagem de 77,22%, pois os gestores ainda não aumentaram os esforços na prevenção de riscos, como detalhado na característica Gestão de Riscos, a qual permaneceu atingindo 40% da meta estabelecida nas duas avaliações. Por fim, os outros indicadores, Criação de Nicho e Sustentabilidade tiveram discreta evolução como pode ser visto no gráfico.

Com a análise das avaliações das métricas, características e indicadores, chegamos à avaliação da saúde de forma geral, a qual evoluiu de 77,82% para 88,74%. Identificamos, após a análise da segunda avaliação, que o EDGov conseguiu se tornar mais saudável a partir do panorama fornecido na primeira avaliação. Em resumo, após a primeira avaliação, os resultados das métricas foram interpretados e as análises foram geradas para mostrar aos gestores quais componentes e aspectos do EDGov precisavam ser priorizados

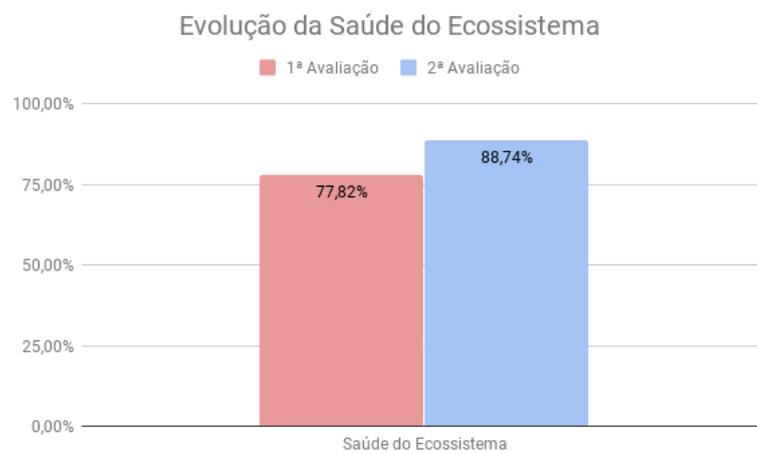
Figura 28 – Avaliação do Estado da Saúde - Indicadores



Fonte: A autora (2019)

para melhorar a saúde do ecossistema. Após o período de 6 meses, outra avaliação foi executada e alguns resultados mostraram melhora e outros permaneceram como estavam, mas não houve detecção de piora. Portanto, chegamos a conclusão que o FASED foi adequado para avaliar a saúde do EDGov pois atingiu o seu objetivo principal de fornecer um panorama da saúde do ED e assim influenciar nas tomadas de decisões do ecossistema. Como melhoria, propomos novas avaliações para manter o monitoramento da saúde do EDGov e assim analisar a evolução dos indicadores, características e métricas. Além disso, o FASED demonstrou ser de fácil aplicação e adaptação a um domínio específico.

Figura 29 – Avaliações da Saúde em Períodos de Tempo Distintos



Fonte: A autora (2019)

5.3 SÍNTESE DO CAPÍTULO

Este Capítulo apresentou um guia de instanciação do FASED. Esse guia foi construído com o intuito de auxiliar os responsáveis pelo ecossistema a avaliar a saúde dos EDs que participam e assim detectar pontos de melhorias. O guia apresenta um processo composto por ações bem definidas, desde a identificação dos gestores e do domínio, até a avaliação utilizando metas e métricas.

Com o objetivo de melhorar o entendimento, nós aplicamos o FASED em um cenário fictício de um Ecossistema de Dados Abertos Governamentais. Nesse cenário nós realizamos duas avaliações da saúde do EDGov a fim de mapear o estado da saúde em períodos de tempo distintos, acompanhar o impacto das análises da primeira avaliação na saúde do EDGov após a segunda avaliação, e destacar os benefícios do uso do *framework* em um cenário próximo à realidade.

No próximo Capítulo será apresentada a avaliação do *framework* proposto neste trabalho.

6 AVALIAÇÃO DO FASED

6.1 GRUPO FOCAL

A avaliação do *framework* FASED foi conduzida por um estudo empírico utilizando o método de grupo focal. De acordo com Kontio, Lehtola e Bragge (2004), um grupo focal é uma técnica que envolve um grupo de participantes reunidos para discutir um determinado problema ou avaliar um determinado tópico. Na literatura, a técnica de grupo focal também é chamada de "Entrevista em Profundidade de Grupo (*Group Depth Interview*)" ou "Entrevista Focada (*Focused Interview*)" Barry, Steyn e Brent (2008). Nos últimos anos, os grupos focais se tornaram populares em diversas áreas como Medicina, Ciências Sociais, Biologia e Ciências da Informação Zaganelli et al. (2015).

Nessa técnica, os participantes irão participar de discussões ou entrevistas organizadas, com um pequeno grupo selecionado de pessoas, discutindo um tópico específico sob a orientação de um facilitador ou moderador Blackburn e Stokes (2000). Chiara (2005) ressalta que o grupo focal não busca obter consenso, o moderador é quem deve criar condições para que diferentes percepções e pontos de vista sejam colocados durante as sessões. Além disso, Chiara (2005) alerta que as discussões inerentes ao processo de adoção dessa técnica devem ocorrer em clima de tranquilidade, sem pressões, de modo que se possa garantir a troca de opiniões em relação ao objeto de estudo.

Para Kontio, Lehtola e Bragge (2004) o grupo focal é um método rápido e econômico para obter experiências de profissionais e usuários, pois pode fornecer dados qualitativos e informações ricas acerca do tema discutido. Além disso, com a aplicação desse método é possível revelar percepções que são difíceis ou caras de capturar com outros métodos Kontio, Lehtola e Bragge (2004). Dessa forma, por possuir uma característica mais dinâmica, os dados gerados pelos grupos focais tendem a ser mais ricos e profundos do que aqueles coletados por métodos de coleta um-para-um (*one-to-one*), como entrevistas ou pesquisas Barry, Steyn e Brent (2008).

6.2 ORGANIZAÇÃO DO GRUPO FOCAL

A organização do Grupo Focal seguiu as etapas sugeridas por Chiara (2005). Para a execução do grupo focal foram definidas três fases: planejamento, condução do grupo focal e análise dos dados. Cada etapa será detalhada nas próximas seções.

6.2.1 Planejamento

Seguindo as recomendações de Chiara (2005), os primeiros passos para execução do grupo focal foram: definir os objetivos do estudo, estipular os critérios de seleção dos partici-

pantes, decidir o local de realização do grupo focal, a duração da sessão, as questões que seriam realizadas e elaborar os documentos que seriam entregues aos participantes com os objetivos do estudo.

O objetivo principal do estudo foi realizar uma avaliação de alto nível a respeito da viabilidade, completude e adequação do FASED. Para essa avaliação foram considerados dez possíveis participantes que poderiam atender aos critérios estipulados. Esses critérios foram:

- conhecimento e expertise em Ecossistemas de Dados;
- conhecimento em Saúde de Ecossistemas de qualquer domínio;
- disposição em compartilhar suas experiências e opiniões.

Após a definição desses critérios, foi enviado um convite para os dez possíveis profissionais que, ao nosso ver, se adequariam aos critérios estipulado. Porém, conseguir conciliar suas agendas foi algo bem complicado, visto que esses profissionais geralmente são muito ocupados. Dessa forma, após muitas discussões e propostas de datas e horários conseguimos fechar em oito profissionais que se dispuseram a participar da pesquisa.

Na Tabela 10 são apresentadas informações gerais dos participantes deste estudo. Todos os participantes são residentes do estado de Pernambuco, Brasil. Quanto aos seus cargos, dois trabalham como Analista de Sistema, um como Engenheiro de Dados, um como Pesquisador e outro como Técnico de Tecnologia da Informação. Temos também um Professor e um Engenheiro de Software. Todos são estudantes (mestrado e doutorado). Dentre os oito, três deles desenvolvem atividades na área de ecossistema de dados apenas no meio acadêmico. Enquanto os outros cinco desenvolvem no meio acadêmico e profissional. Em relação a formação, cinco possuem graduação, um especialização e dois mestrado. Além disso, cinco dos oito participantes tem, pelo menos, dois anos de experiência na área de ecossistemas de dados.

O grupo focal foi realizado no Centro de Informática da Universidade Federal de Pernambuco - CIn/UFPE. O local escolhido foi de comum acordo para todos os participantes. A sessão ocorreu no mês de Julho de 2019, com início as 9h00 e término as 11h30, ou seja, foram 2:30h 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, que ficou responsável por realizar todas as anotações pertinentes durante a sessão e pela gravação de áudio das discussões.

Tabela 10 – Informações sobre os Participantes do Grupo Focal

Participante	Função/ Cargo	Formação	Experiência com Ecos. de Dados	Meios que desenvolve as atividades
--------------	---------------	----------	--------------------------------	------------------------------------

Participante 1	Estudante de Mestrado/ Técnico de TI	Especialização	Entre 2 e 3 anos	Acadêmico e Profissional
Participante 2	Estudante de Mestrado/ Engenheiro de Dados	Graduação	Entre 2 e 3 anos	Acadêmico e Profissional
Participante 3	Estudante de Mestrado/ Pesquisador	Graduação	Entre 2 e 3 anos	Acadêmico
Participante 4	Estudante de Mestrado/ Engenheiro de Software	Graduação	Não possui experiência	Acadêmico e Profissional
Participante 5	Estudante de Mestrado/ Analista de Dados	Graduação	Entre 2 e 3 anos	Acadêmico e Profissional
Participante 6	Estudante de Doutorado/ Professor	Mestrado	Não possui experiência	Acadêmico
Participante 7	Estudante de Doutorado/ Analista de Sistemas	Mestrado	Entre 4 e 5 anos	Acadêmico e Profissional
Participante 8	Estudante de Mestrado	Graduação	Menos de 2 anos	Acadêmico

Fonte: A autora (2019)

6.2.2 Condução do Grupo Focal

A sessão do grupo focal iniciou com uma breve apresentação do moderador, observador e participantes. Cada um teve a oportunidade de dizer seu nome, o que faz e sua experiência na área. Após isso, foi entregue a cada participante o termo de consentimento para que o participante ficasse a par 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, 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 FASED.

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. Além disso, também foram apresentadas algumas regras de funcionamento. Essas regras foram:

- Nós queremos ouvir a opinião de todos;
- Todos os comentários são válidos;
- Não existem respostas corretas ou incorretas;
- Não fugir do 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 dezoito perguntas a respeito dos indicadores, características e métricas do FASED. Essas

perguntas eram compostas por *check-box*, possibilitando, dessa forma, que os participantes selecionassem mais de uma resposta para cada pergunta. Por último, na terceira fase seriam quatro perguntas gerais, a fim de coletar dados de avaliação de qualidade, bem como comentários gerais. O formulário aplicado pode ser encontrado em: <<http://bit.ly/avaliacaoFASED>>

6.2.3 Análise dos Dados

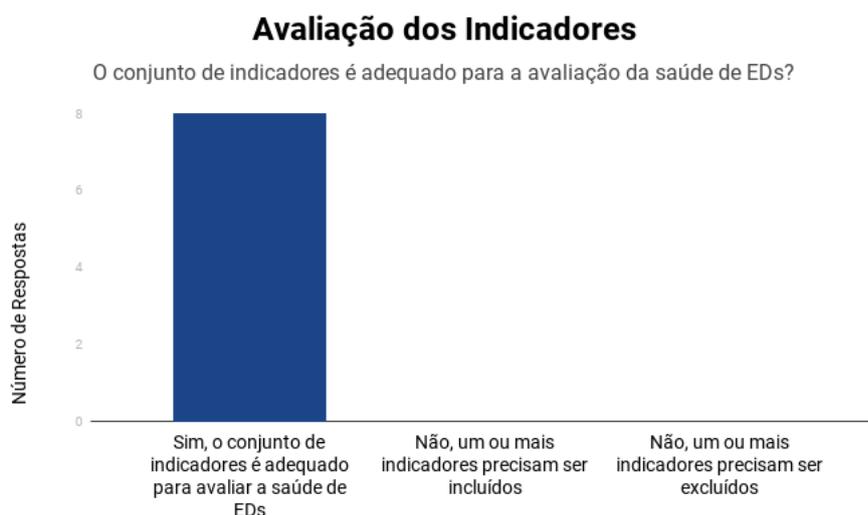
Esta seção apresenta a análise dos dados coletados no grupo focal, apresentando cada pergunta realizada, as respostas dos participantes, as sugestões apontadas por eles e a análise e justificativa para absorver ou não as sugestões no trabalho. Para cada seção da Fase 2 do grupo focal foram apresentadas 2 perguntas, a primeira se o elemento em questão era adequado para avaliar a saúde de EDs, e a segunda se a sua descrição era clara, facilitando o entendimento e aplicabilidade no contexto de avaliação da saúde de EDs.

6.2.3.1 Avaliação dos Indicadores

Pergunta 1

A primeira pergunta foi relacionada ao conjunto de indicadores definidos no FASED. Foi perguntado se o conjunto de indicadores é adequado, viável e completo para avaliar a saúde de EDs. Além disso, com essa pergunta buscamos identificar lacunas, erros e/ou melhorias. A Figura 30 resume as respostas.

Figura 30 – Resultado da Avaliação dos Indicadores - Pergunta 1



Fonte: A autora (2019)

Todos os participantes selecionaram a opção "Sim, o conjunto de indicadores é adequado para avaliar a saúde de EDs".

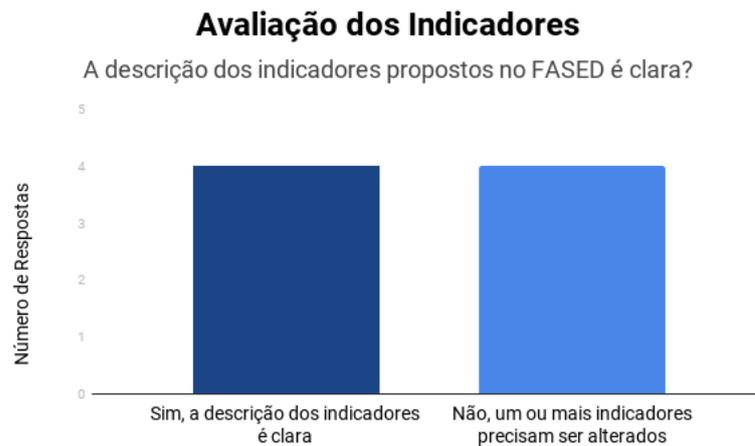
Participante 4 comentou que "Acredito que sejam suficientes para avaliar a saúde e não precisa de mais algum".

Participante 8 complementou dizendo que "Fiquei com dúvida no indicador de Produtividade, ao falar em "Atividade" se está incluso avaliar e indicar os relacionamentos entre os atores." Quanto ao comentário do *Participante 8*, o termo "Atividade" se refere a atividades que estimulem direta ou indiretamente a produção e consumo de recursos, e que possam ser medidos de forma quantitativa. Métricas qualitativas de saúde não são cobertas pelo FASED, para tanto se faz necessário um estudo mais aprofundado e isso é proposto como trabalho futuro.

Pergunta 2

A segunda pergunta foi relacionada à descrição dos indicadores, se eles estão descritos de forma clara. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 31 resume as respostas.

Figura 31 – Resultado da Avaliação dos Indicadores - Pergunta 2



Fonte: A autora (2019)

Participante 1 selecionou "Não, um ou mais indicadores precisam ser alterados" e sugeriu "Esclarecer melhor a diferença entre Robustez e Sustentabilidade, pois o termo sobrevivência foi citado em ambas as definições".

Os *Participantes 3, 4 e 7* selecionaram "Não, um ou mais indicadores precisam ser alterados" e comentaram que o termo "Robustez" não está alinhado com a descrição apresentada, pois a descrição fala da capacidade de permanecer estável e cita o termo "sobrevivência". Sugeriram também a troca do nome "Robustez" por "Estabilidade" ou "Manutenibilidade".

Participante 2 selecionou "Sim, a descrição dos indicadores está clara" e escreveu no formulário "O texto de descrição dos indicadores auxiliam na interpretação. O indicador Robustez e Sustentabilidade podem causar uma confusão no início, devido ao termo "so-

brevivência". Apesar disso, com o decorrer do texto, as características de cada indicador fica mais evidente, facilitando a interpretação."

Participante 5 selecionou "Sim, a descrição dos indicadores está clara" e escreveu no formulário "O indicador Robustez poderia ter seu nome alterado para algo que descrevesse, ou chegasse mais próximo do objetivo proposto. Isto é, sua descrição faz referência, no meu entendimento, a coisas relacionadas a estabilidade do ED, no entanto seu nome sugere que seja algo de viés mais técnico, como por exemplo, a capacidade de processamento e amplitude do ED."

Participante 6 selecionou "Sim, a descrição dos indicadores está clara" e escreveu no formulário "Avaliação da saúde de ecossistemas de maneira geral utilizam o termo "Robustez" e ele é um pouco questionado. Seria interessante frisar a relação dele com estabilidade e manutenibilidade."

Os demais participantes selecionaram "Sim, a descrição dos indicadores está clara" e não deixaram comentário no questionário.

Em relação aos comentários dos *Participantes 1 e 2*, nós concordamos que o uso de um mesmo termo nas definições de "Robustez" e "Sustentabilidade" poderia deixar o leitor em dúvida. Por isso, reformulamos as descrições e deixamos mais clara a diferença entre os dois.

Os comentários dos *Participantes 3, 4, 5, 6 e 7* sugeriram a mudança do nome do indicador "Robustez" ou a melhoria da sua definição. Nós concordamos que a descrição desse indicador poderia ser mais detalhada e executamos essa melhoria frisando a sua relação com a capacidade de permanecer estável. Entretanto, não concordamos com a mudança do nome, pois "Robustez" é um indicador amplamente difundido e consolidado na literatura de avaliação da saúde de Ecossistemas de Negócios e *Software*, e é utilizado com o mesmo objetivo no contexto de EDs para avaliar a saúde.

Com isso, atribuímos a esses questionamentos o porque de 50% dos participantes terem sugerido que os indicadores precisavam de uma descrição mais clara.

6.2.3.2 Avaliação das Características do Indicador de Produtividade

Pergunta 1

A primeira pergunta foi relacionada às características do indicador Produtividade definidas no FASED. Foi perguntado se o conjunto de características do indicador Produtividade é adequado, viável e completo para avaliar a saúde de EDs. Além disso, com essa pergunta buscamos identificar lacunas, erros e/ou melhorias. A Figura 32 resume as respostas.

Todos os participantes selecionaram a resposta "Sim, o conjunto de características é adequado para avaliar a saúde de EDs" e o *Participante 4* comentou "O conjunto atende, não precisa inclusão de mais características pra esse indicador".

Pergunta 2

Figura 32 – Resultado da Avaliação das Características de Produtividade - Pergunta 1



Fonte: A autora (2019)

A segunda pergunta foi relacionada à descrição das características, se elas estão descritas de forma clara. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 33 resume as respostas.

Figura 33 – Resultado da Avaliação das Características de Produtividade - Pergunta 2



Fonte: A autora (2019)

Os *Participantes 2, 3 e 5* selecionaram a resposta "Não, um ou mais características precisam ser alteradas" e sugeriram adição de exemplos na característica "Atividade" para deixar mais claro que tipos de atividades ela quer medir.

Os demais participantes selecionaram "Sim, a descrição das características está clara" e não deixaram comentário no questionário.

Em relação às sugestões propostas pelos *Participantes 2, 3 e 5*, nós concordamos e adicionamos alguns exemplos de atividades que podem ser consideradas nessa característica,

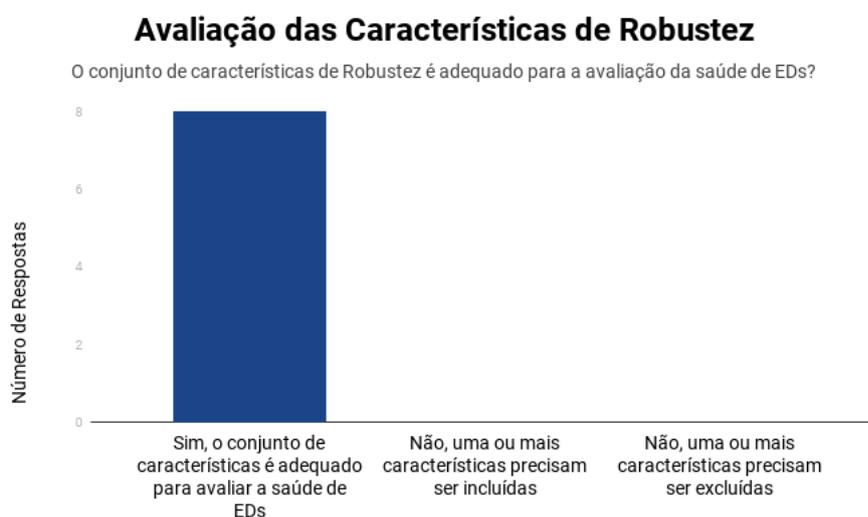
como *workshops*, palestras, *hackathons*, entre outros.

6.2.3.3 Avaliação das Características do Indicador de Robustez

Pergunta 1

A primeira pergunta foi relacionada às características do indicador Robustez definidas no FASED. Foi perguntado se o conjunto de características do indicador Robustez é adequado, viável e completo para avaliar a saúde de EDs. Além disso, com essa pergunta buscamos identificar lacunas, erros e/ou melhorias. A Figura 34 resume as respostas.

Figura 34 – Resultado da Avaliação das Características de Robustez - Pergunta 1



Fonte: A autora (2019)

Todos os participantes selecionaram a resposta "Sim, o conjunto de características é adequado para avaliar a saúde de EDs" e não deixaram comentário no questionário.

Pergunta 2

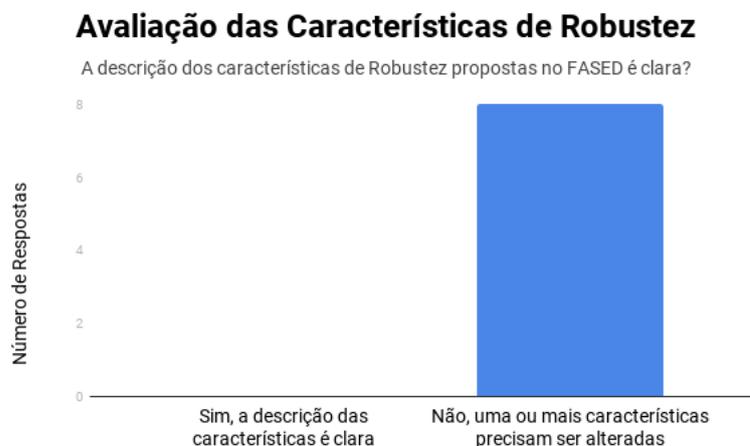
A segunda pergunta foi relacionada à descrição das características, se elas estão descritas de forma clara. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 35 resume as respostas.

Todos os participantes selecionaram a resposta "Não, um ou mais características precisam ser alteradas".

Participante 1 comentou "Não ficou claro para mim se o foco do indicador seria financeiro, tecnológico, ou uma combinação de ambos".

Os *Participantes 2, 3, 4, 5, 6, 7 e 8* sugeriram a troca do termo "Persistência" na característica de "Persistência dos dados", pois o termo remete ao termo técnico "persistência em Banco de Dados" e pode confundir o leitor. Eles comentaram que uma alternativa é trocar o termo, a fim de evitar essa chance de má interpretação ou adicionar exemplos durante a descrição se permanecer esse termo.

Figura 35 – Resultado da Avaliação das Características de Robustez - Pergunta 2



Fonte: A autora (2019)

Participante 8 comentou "A característica de gestão como é focada na gestão preventiva, poderia ser renomeada para "Gestão de Riscos".

Em relação ao comentário do *Participante 1*, é possível agrupar características de diferentes áreas, desde que todas estejam relacionadas com o objetivo principal do indicador, que nesse caso é avaliar a capacidade do ED em permanecer estável ao enfrentar disrupções.

Já em relação à sugestão dos *Participantes 2, 3, 4, 5, 6, 7 e 8* para a troca do termo "Persistência" na característica de "Persistência dos dados", nós concordamos em alguns pontos. O termo "Persistência dos dados" não está incorreto, pois o objetivo inicial dessa característica era prover mecanismos que garantissem que os dados não seriam perdidos em situações inesperadas. Entretanto, percebemos que seria mais interessante abranger mais tipos de recursos e que avaliar somente a persistência dos dados a limitaria. Por isso, aumentamos o escopo para outros tipos de recursos, como aplicações e serviços, e trocamos o termo para "Garantia de Disponibilidade". Esse novo nome permanece alinhado com o objetivo da característica de representar a existência de mecanismos que garantam a disponibilidade contínua dos recursos mesmo em situações de instabilidade.

Nós concordamos com a sugestão do *Participante 8*, sobre adicionar o termo "Riscos" no nome da característica de gestão, pois deixa mais alinhado com o objetivo da característica.

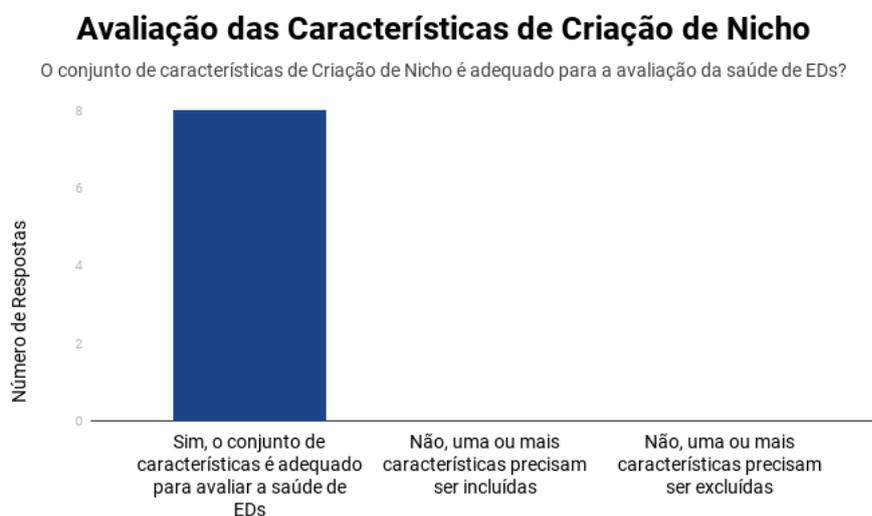
6.2.3.4 Avaliação das Características do Indicador de Criação de Nicho

Pergunta 1

A primeira pergunta foi relacionada às características do indicador Criação de Nicho definidas no FASED. Foi perguntado se o conjunto de características do indicador Criação de Nicho é adequado, viável e completo para avaliar a saúde de EDs. Além disso, com

essa pergunta buscamos identificar lacunas, erros e/ou melhorias. A Figura 36 resume as respostas.

Figura 36 – Resultado da Avaliação das Características de Criação de Nicho - Pergunta 1



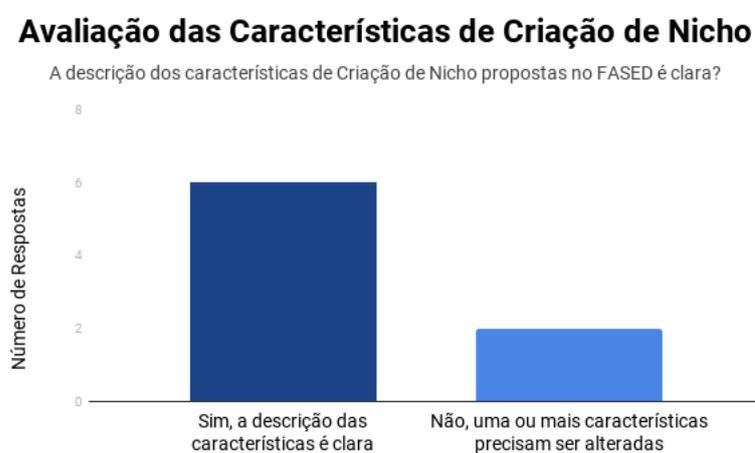
Fonte: A autora (2019)

Todos os participantes selecionaram a resposta "Sim, o conjunto de características é adequado para avaliar a saúde de EDs" e não deixaram comentário no questionário.

Pergunta 2

A segunda pergunta foi relacionada à descrição das características, se elas estão descritas de forma clara. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 37 resume as respostas.

Figura 37 – Resultado da Avaliação das Características de Criação de Nicho - Pergunta 2



Fonte: A autora (2019)

Participante 3 selecionou a resposta "Não, um ou mais características precisam ser alteradas" e comentou "*Talvez mudar o termo "variedade". Ele está muito abrangente*".

Participante 7 selecionou a resposta "Não, um ou mais características precisam ser alteradas" e sugeriu "Acredito que no indicador de "Criação de Nicho" deveria ter somente a característica de "Variedade". A característica de "Diversidade Financeira" poderia estar junto de "Consistência Financeira" no indicador de "Robustez".".

Os demais participantes selecionaram "Sim, a descrição das características está clara" e não deixaram comentário no questionário.

Em relação ao comentário do *Participante 3* nós discordamos, pois o objetivo é que o termo seja amplo e englobe diferentes tipos de elementos que gerem nichos no Ecossistema. Nós concordamos com o comentário do *Participante 7*, pois após analisar a gama de resultados que podem ser coletados com essa característica, identificamos que a diversidade financeira que queremos medir está alinhada com o objetivo do indicador Robustez. A diversidade financeira influencia a Robustez ao prover diferentes formas de monetização e estruturas organizacionais que auxiliam na consistência financeira e na gestão de riscos, respectivamente. O indicador de Criação de Nicho foi alterado para ser composto somente pela característica "Variedade", a qual indica a existência de nichos analisando a variedade dos elementos do ecossistema, removemos a característica "Diversidade Financeira" e alocamos as métricas dessa característica para o indicador de Robustez, o qual chegamos a conclusão ser mais adequado.

6.2.3.5 Avaliação das Características do Indicador de Sustentabilidade

Pergunta 1

A primeira pergunta foi relacionada às características do indicador Sustentabilidade definidas no FASED. Foi perguntado se o conjunto de características do indicador Sustentabilidade é adequado, viável e completo para avaliar a saúde de EDs. Além disso, com essa pergunta buscamos identificar lacunas, erros e/ou melhorias. A Figura 38 resume as respostas.

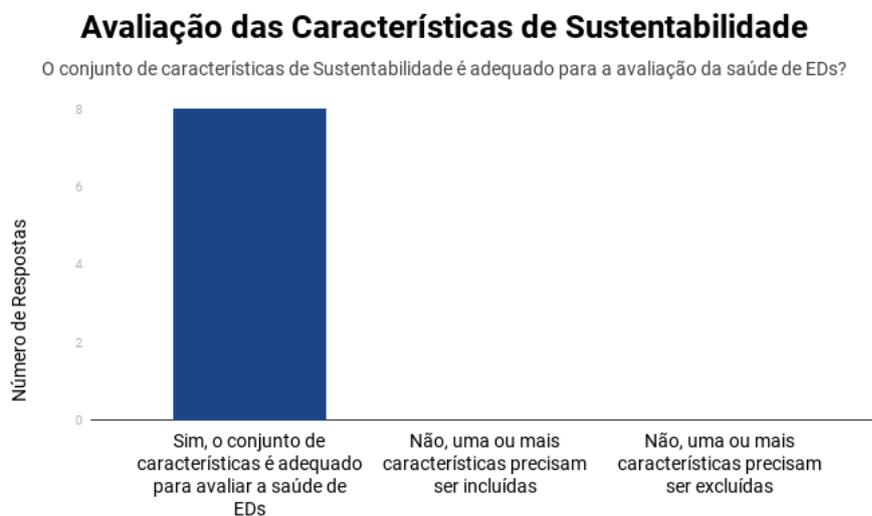
Todos os participantes selecionaram a resposta "Sim, o conjunto de características é adequado para avaliar a saúde de EDs" e não deixaram comentário no questionário.

Pergunta 2

A segunda pergunta foi relacionada à descrição das características, se elas estão descritas de forma clara. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 39 resume as respostas.

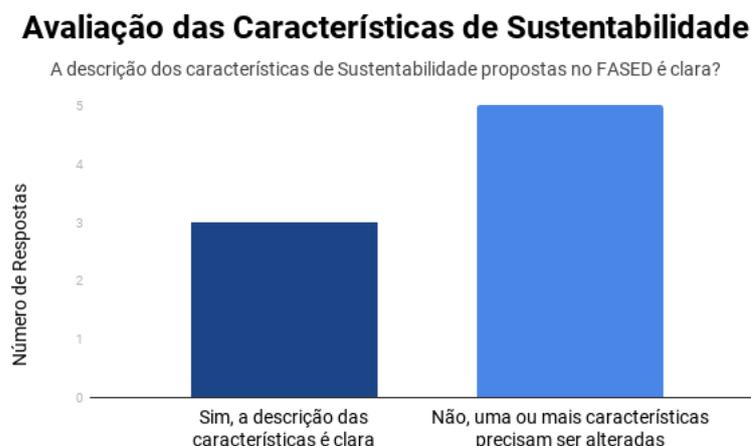
Os *Participantes 1, 4, 5, 6 e 8* selecionaram a resposta "Não, um ou mais características precisam ser alteradas" e comentaram que a característica de "Habilidade de Regeneração" poderia ser melhor explicada, com exemplos, ou modificar o nome da característica. *Participante 5* afirmou "A característica Habilidade de Regeneração poderia ter o termo "regeneração" modificado. Pois algo que se regenera um dia deixou de existir.". Sobre esse comentário, argumentamos que o intuito dessa característica é analisar a evolução da rede de atores, se apoiando em números e em períodos de tempo, onde houve

Figura 38 – Resultado da Avaliação das Características de Sustentabilidade - Pergunta 1



Fonte: A autora (2019)

Figura 39 – Resultado da Avaliação das Características de Sustentabilidade - Pergunta 2



Fonte: A autora (2019)

maiores entradas e/ou saídas de atores, para detectar se o ED já precisou se regenerar, e sim, indicando que essa característica é presente no ED. Pois, um ecossistema que já precisou se regenerar, tem mais chances de sobreviver no futuro. Nós optamos por detalhar mais essa característica, pois ela é amplamente referenciada na literatura e é alinhada com os objetivos do indicador de Sustentabilidade.

Os demais participantes selecionaram "Sim, a descrição das características está clara" e não deixaram comentário no questionário.

6.2.3.6 Avaliação das Métricas do Indicador de Produtividade

Pergunta 1

A primeira pergunta foi relacionada às métricas do indicador Produtividade definidas no FASED. Foi perguntado se o conjunto de métricas do indicador Produtividade é adequado para avaliar a saúde de EDs. Além disso, com essa pergunta buscamos identificar erros e/ou melhorias. A Figura 40 resume as respostas.

Figura 40 – Resultado da Avaliação das Métricas de Produtividade - Pergunta 1



Fonte: A autora (2019)

Todos os participantes selecionaram a resposta "Sim, o conjunto de métricas é adequado para avaliar a saúde de EDs" e não deixaram comentário no questionário.

Pergunta 2

A segunda pergunta foi relacionada à descrição das métricas, se o nível de detalhamento facilita seu entendimento e auxilia na aplicação delas. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 41 resume as respostas.

Participante 6 selecionou a resposta "Não, uma ou mais informações das métricas precisam ser alteradas" e comentou "A métrica que pergunta sobre a existência de materiais de referência poderia ser substituída pela quantificação desses materiais. Saber se existe é válido, mas se existir, saber a quantidade não seria interessante?".

Os demais participantes selecionaram "Sim, as métricas são de fácil entendimento e aplicabilidade" e não deixaram comentário no questionário.

Nós concordamos com a sugestão de melhoria do *Participante 6* e alteramos a métrica, pois é mais relevante ter essa métrica quantificando os materiais do que indicando somente a existência deles. O impacto de ter muitos materiais de referência é diferente de ter apenas um, pois influencia na produtividade dos atores que utilizam esses materiais para produzirem e consumirem recursos.

Figura 41 – Resultado da Avaliação das Métricas de Produtividade - Pergunta 2



Fonte: A autora (2019)

6.2.3.7 Avaliação das Métricas do Indicador de Robustez

Pergunta 1

A primeira pergunta foi relacionada às métricas do indicador Robustez definidas no FASED. Foi perguntado se o conjunto de métricas do indicador Robustez é adequado para avaliar a saúde de EDs. Além disso, com essa pergunta buscamos identificar erros e/ou melhorias. A Figura 42 resume as respostas.

Figura 42 – Resultado da Avaliação das Métricas de Robustez - Pergunta 1



Fonte: A autora (2019)

Participantes 4, 5 e 6 selecionaram a resposta "Não, uma ou mais métricas precisam ser incluídas".

Participantes 4 e 5 sugeriram a inclusão de uma nova métrica que verifique a existência de documentação detalhada sobre os processos de produção e consumo do ecossistema. Nós concordamos em adicionar essa métrica pois existir a documentação dos processos é mais uma forma de apoiar a execução das atividades do ecossistema no dia a dia e em situações de instabilidade.

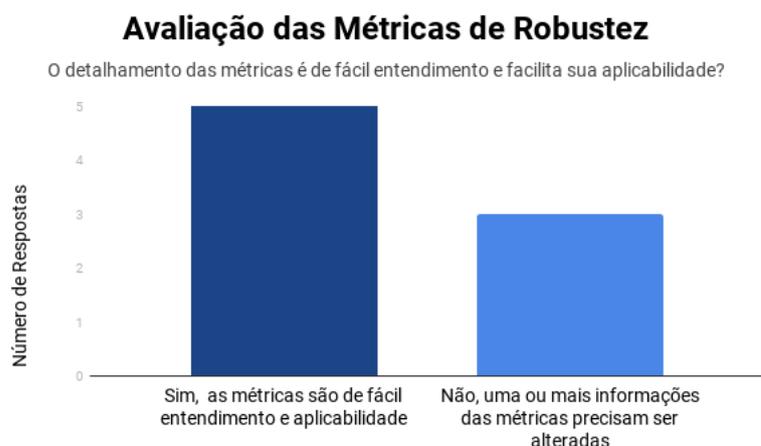
Participante 6 comentou "*Para as métricas que perguntam sobre a existência de investidores públicos, privados, pontuais e permanentes, sugiro incluir métricas que calculem o número de investidores em função do tempo, pois isso agrega mais valor na análise dos resultados.*". Nós concordamos com esse comentário e incluímos novas métricas para calcular a quantidade de investidores ao longo do tempo. Saber o número de investidores em função do tempo auxilia na tomada de decisões, pois revela atores com potencial de se tornarem investidores permanentes, por exemplo. Somado a isso, torna-se possível mapear a existência e evolução do número de investidores ao longo do tempo. Portanto, após a inclusão dessas novas métricas, deixou de ser útil saber a existência desses investidores, já que as novas métricas já trazem essa informação.

Os demais participantes selecionaram "Sim, o conjunto de métricas é adequado para avaliar a saúde de EDs" e não deixaram comentário no questionário.

Pergunta 2

A segunda pergunta foi relacionada à descrição das métricas, se o nível de detalhamento facilita seu entendimento e auxilia na aplicação delas. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 43 resume as respostas.

Figura 43 – Resultado da Avaliação das Métricas de Robustez - Pergunta 2



Fonte: A autora (2019)

Participantes 1 e 4 selecionaram a resposta "Não, uma ou mais informações das métricas precisam ser alteradas". O *Participante 1* comentou "*Em relação à métrica de múltiplos mecanismos de acesso, talvez seja necessário mudar a descrição da métrica para deixar*

mais claro do que se trata esses "mecanismos de acesso". E o *Participante 4* comentou "Agrega mais valor se nas métricas sobre a existência de investidores públicos, privados, pontuais e permanentes, fosse colocado a quantidade de investidores em função do tempo. Não só a binaridade se existe ou não".

Os demais participantes selecionaram "Sim, as métricas são de fácil entendimento e aplicabilidade" e não deixaram comentário no questionário.

Com relação ao comentário do *Participante 1*, a métrica já tem a palavra "acesso" tanto no título quanto na descrição da métrica. Para não gerar dúvidas no leitor, incluímos mais detalhes na descrição da métrica. Já em relação à sugestão do *Participante 4*, nós incluímos novas métricas, como foi dito na análise dos resultados da pergunta 1 desta seção. Sobre as métricas que mediam a existência de investidores, elas foram excluídas, pois os resultados das novas métricas são mais completos e já trazem essa informação.

6.2.3.8 Avaliação das Métricas do Indicador de Criação de Nicho

Pergunta 1

A primeira pergunta foi relacionada às métricas do indicador Criação de Nicho definidas no FASED. Foi perguntado se o conjunto de métricas do indicador Criação de Nicho é adequado para avaliar a saúde de EDs. Além disso, com essa pergunta buscamos identificar erros e/ou melhorias. A Figura 44 resume as respostas.

Figura 44 – Resultado da Avaliação das Métricas de Criação de Nicho - Pergunta 1



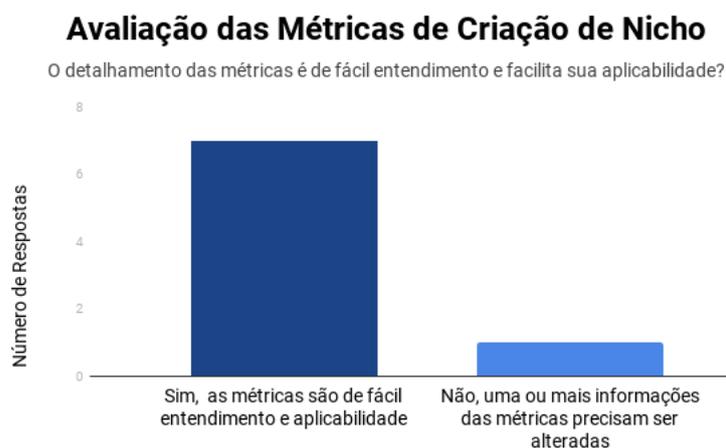
Fonte: A autora (2019)

Todos os participantes selecionaram a resposta "Sim, o conjunto de métricas é adequado para avaliar a saúde de EDs" e não deixaram comentário no questionário.

Pergunta 2

A segunda pergunta foi relacionada à descrição das métricas, se o nível de detalhamento facilita seu entendimento e auxilia na aplicação delas. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 45 resume as respostas.

Figura 45 – Resultado da Avaliação das Métricas de Criação de Nicho - Pergunta 2



Fonte: A autora (2019)

Participante 1 selecionou a resposta "Não, uma ou mais informações das métricas precisam ser alteradas". O *Participante 1* comentou "A métrica do número de tipos de papéis, talvez fosse interessante modificar para refletir o número de pessoas por papel". Nós não concordamos em modificar essa métrica, pois a variedade de papéis tem a capacidade de formar nichos entre diferentes tipos de atores com o objetivo de gerar novas redes ou recursos, por exemplo. A sugestão de criar uma métrica com o número de atores por papel está alinhada com o objetivo de medir o equilíbrio de esforços dentro do ecossistema, que é uma característica do indicador de Sustentabilidade. Portanto, absorvemos essa sugestão e adicionamos uma nova métrica nesse outro indicador.

Os demais participantes selecionaram "Sim, as métricas são de fácil entendimento e aplicabilidade" e não deixaram comentário no questionário.

6.2.3.9 Avaliação das Métricas do Indicador de Sustentabilidade

Pergunta 1

A primeira pergunta foi relacionada às métricas do indicador Sustentabilidade definidas no FASED. Foi perguntado se o conjunto de métricas do indicador Sustentabilidade é adequado para avaliar a saúde de EDs. Além disso, com essa pergunta buscamos identificar erros e/ou melhorias. A Figura 46 resume as respostas.

Participante 5 selecionou a resposta "Não, um ou mais métricas precisam ser incluídas". Ele escreveu no formulário "Adicionar uma nova métrica: Atores participantes ativos desde o início do ED.". Conforme analisamos, já tínhamos uma métrica que media a razão

Figura 46 – Resultado da Avaliação das Métricas de Sustentabilidade - Pergunta 1



Fonte: A autora (2019)

entre o número de atores desde o início do ecossistema e o total de atores do ED, que tinha o mesmo objetivo mas faltava detalhar melhor. Então ao invés de adicionar uma nova métrica, nós atualizamos a descrição da métrica existente incluindo o termo "ativo" para ficar mais claro.

Os demais participantes selecionaram a resposta "Sim, o conjunto de métricas é adequado para avaliar a saúde de EDs" e não deixaram comentário no questionário.

Pergunta 2

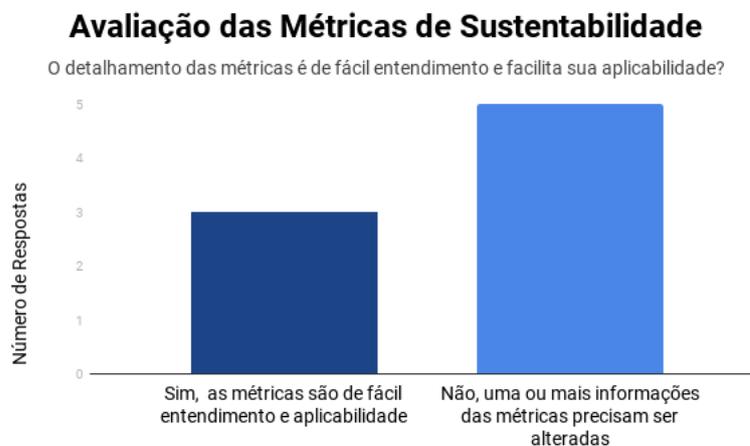
A segunda pergunta foi relacionada à descrição das métricas, se o nível de detalhamento facilita seu entendimento e auxilia na aplicação delas. Além disso, com essa pergunta buscamos identificar erros e/ou possibilidades de melhoria. A Figura 47 resume as respostas.

Participantes 1, 3, 4, 5 e 6 selecionaram a resposta "Não, uma ou mais informações das métricas precisam ser alteradas".

Participante 1 escreveu no formulário "Refletir sobre a necessidade de incluir uma nova métrica, ou mudar o texto da métrica que calcula o número de trabalhos científicos sobre o ED, para incluir citações do ED na literatura cinza.". Consideramos relevante essa sugestão e alteramos o termo "trabalhos científicos" para "literatura", pois abrange mais tipos de documentos, como patentes, normas, documentações comerciais, técnicas, entre outros. Portanto, dependendo do domínio a ser aplicado, a métrica pode ser especializada para qualquer tipo de documento da literatura.

Participante 3 comentou "Na métrica que mede a existência de remoção de dados inconsistentes, trocar "remoção" por "refinamento". Nem sempre o dado inconsistente precisa ser removido. Em alguns casos, a adição de novos dados ou metadados pode eliminar a inconsistência.". Nós concordamos e alteramos a métrica, pois de acordo com a litera-

Figura 47 – Resultado da Avaliação das Métricas de Sustentabilidade - Pergunta 2



Fonte: A autora (2019)

tura, refinar os dados significa enriquecer e/ou limpar os dados Santos (2018), e isso está alinhado com o objetivo de eliminar inconsistências.

Participantes 4 e 5 sugeriram abranger a métrica que calcula o número de mensagens em mídias sociais, "para englobar a publicação de materiais em um contexto mais amplo.". Nós concordamos e alteramos de "mensagens" para "citações" a fim de propor uma métrica com espectro mais amplo e, quando necessário, especializar essa métrica para mensagens em redes sociais, posts em blogs, ou outros tipos de forma de citação, dependendo do contexto que o *framework* estiver sendo aplicado.

Participante 6 sugeriu alterar o cálculo das métricas nomeadas por "Centralização de Esforços" e "Descentralização de Parcerias", para ao invés de obter respostas sobre a existência desses aspectos, calcular a razão entre o número de parcerias/atividades e a quantidade total de parcerias/atividades. Ao analisar essa sugestão nós percebemos que essa mudança traz informações mais relevantes sobre as características a qual elas pertencem, como a porcentagem da centralização dos esforços e a porcentagem da distribuição de parcerias em outros países/estados/cidades. Por isso, alteramos a fórmula para calcular essas métricas.

Participantes 4 e 6 comentaram que o uso do termo "Facilidade de Acesso", na métrica que pergunta se os dados são de fácil acesso, deveria ser alterado porque a resposta para essa métrica é subjetiva, o que para uns é fácil, para outros pode não ser. Argumentamos que a métrica "Facilidade de Acesso" tem o objetivo de avaliar se existe meios que facilitem o acesso aos dados, como a existência de API's, portais para *download*, entre outros. Com isso, adicionamos na descrição da métrica exemplos de meio de acesso para deixar mais claro para o leitor.

Os demais participantes selecionaram "Sim, as métricas são de fácil entendimento e aplicabilidade" e não deixaram comentário no questionário.

6.2.3.10 Perguntas Gerais

A última fase do grupo focal foi composta por quatro perguntas discursivas. Nelas pedimos aos participantes numa escala de 1 a 5, na qual 1 significa "Discordo Fortemente" e 5 significa "Concordo Fortemente", como eles avaliavam as seguintes questões:

1. O FASED é relevante para avaliar a saúde de Ecossistemas de Dados?
2. O FASED é genérico o suficiente para avaliar a saúde de Ecossistemas de Dados de qualquer domínio?
3. O FASED é adaptável para qualquer domínio de Ecossistemas de Dados?
4. A hierarquia de elementos do FASED é coerente?

A resposta de cada participante a essas perguntas foram apresentadas na Tabela 11.

Tabela 11 – Resultado das Perguntas Gerais sobre o FASED

Participante	FASED é relevante	FASED é genérico	FASED é adaptável	Hierarquia de elementos é coerente
Participante 1	5	5	5	5
Participante 2	5	4	4	5
Participante 3	5	5	5	5
Participante 4	5	5	5	5
Participante 5	5	4	5	5
Participante 6	4	5	5	5
Participante 7	5	5	5	5
Participante 8	5	5	5	5
Média	4.875	4.75	4.875	5

Fonte: A autora (2019)

Participante 2 comentou nas perguntas "FASED é genérico" e "FASED é adaptável" que "A adaptação de algumas métricas (unidades de tempo e etc) e termos podem auxiliar numa maior generalidade. Apenas esse detalhe para atingir o 5."

Participante 4 comentou "Bem relevante. Ainda mais porque não existem outros trabalhos relacionados à saúde de ecossistemas de dados. Pensei em vários cenários de uso como por exemplo: blockchain, distribuição de códigos fonte, dados da saúde pública, etc. Com a hierarquia de indicadores, características e métrica ficou bem fácil de entender e discutir."

Participante 5 fez o seguinte comentário na pergunta "FASED é relevante": "Extremamente importante, compreensível e de utilidade. Ecossistemas de dados podem ser muito grandes a depender do contexto, ter algo que direcione para sua criação, expansão e manutenção, fortalece o desenvolvimento de soluções robustas e confiáveis.". Como resposta para a questão "FASED é genérico", ele afirmou "Algumas métricas são muitos específicas,

como por exemplo, publicação de trabalhos científicos. A depender do domínio, isso não se aplica. Se é para ser um framework genérico, todas as métricas devem ser descritas de forma ampla."

Participante 6 comentou "Importante para gestores fazerem análises sobre a saúde do ecossistema."

Participante 7 comentou "A sugestão de tratar a parte financeira como um indicador a parte, talvez faça mais sentido ao ler os indicadores de robustez e criação de nicho." Nós consideramos que os aspectos financeiros mapeados no indicador de Robustez fornece indícios suficientes sobre a saúde financeira do ecossistema. Essa sugestão pode ser considerada como trabalho futuro por meio de pesquisas mais aprofundadas sobre os aspectos financeiros específicos dos EDs.

Participante 8 adicionou "Acredito que o framework, de maneira geral, está bem amarrado e apresenta muitas questões relevantes para avaliar o Ecossistema de Dados, inclusive fornece parâmetros para uma melhor gestão desse ecossistema auxiliando no processo de decisão."

Os demais participantes não deixaram comentários no questionário.

7 CONCLUSÃO

Neste capítulo descrevemos as considerações finais sobre esta dissertação. Apresentamos as contribuições de pesquisas alcançadas na Seção 7.1, as limitações na Seção 7.2 e as diretrizes para trabalhos futuros na Seção 7.3.

7.1 CONSIDERAÇÕES FINAIS

Neste trabalho, propomos um *Framework* para Avaliação da Saúde de Ecossistemas de Dados, denominado FASED. O *framework* proposto se mostra relevante para o contexto de Ecossistemas de Dados, uma vez que apresenta uma estrutura para guiar e apoiar os gestores na avaliação da saúde de EDs. A avaliação é feita por meio de indicadores, características e métricas que medem aspectos concretos que influenciam direta e indiretamente na estabilidade, sobrevivência, criação de valor e produtividade do ED.

Inicialmente, foi apresentada uma visão geral do que são Ecossistemas de Dados por meio dos detalhes do mapeamento sistemático publicado pelo autor, orientador e co-orientador deste trabalho. Em seguida, apresentamos como a saúde é avaliada nos contextos de Negócios e de *Software*. No Capítulo 3, dissertamos sobre os estudos de saúde em Ecossistemas de Negócios que influenciam nos *frameworks* de avaliação da saúde de Ecossistemas de *Software*, os quais também são explorados. Além disso, apresentamos trabalhos de outras áreas que influenciam o FASED, como modelos de avaliação de qualidade de *software* e os trabalhos que avaliam a qualidade dos dados. Como resultado do mapeamento sistemático na área de EDs e do estudo *ad-hoc* nas áreas mencionadas anteriormente, constatou-se que nenhum trabalho com o propósito de avaliar a saúde de EDs foi encontrado.

Desse modo, propomos um *Framework* para Avaliação da Saúde de Ecossistemas de Dados que tem como objetivo revelar o panorama da saúde de EDs por meio de métricas que avaliam o estado atual da saúde e a sua capacidade em se manter saudável ao longo do tempo. Ademais, ele utiliza indicadores de saúde, como Produtividade, Robustez, Criação de Nicho e Sustentabilidade, para prover uma visão alto nível de toda a saúde do ecossistema, onde cada indicador é constituído por um grupo de características que, por sua vez, reúnem métricas relacionadas. Ele é composto por 4 indicadores, 12 características e 49 métricas. O FASED possui um caráter generalista que proporciona a sua adaptação para qualquer domínio de EDs.

Para avaliarmos o FASED realizamos um grupo focal que contou com 8 participantes da área de Ecossistemas de Dados e de *Software* e que tinham experiência na publicação e/ou consumo de dados. Esse grupo focal foi realizado no Centro de Informática da Universidade Federal de Pernambuco e sua condução foi dividida em três fases. Na primeira

fase, foi solicitado que os participantes respondessem um questionário sobre seus dados profissionais e acadêmicos. Durante a segunda fase houve a aplicação de um questionário contendo 18 perguntas com o objetivo de avaliar os indicadores, características e métricas do FASED. Na terceira e última fase, foram feitas quatro perguntas gerais com o intuito de avaliar se o *framework* é relevante, genérico e adaptável, e se a hierarquia de elementos é coerente.

Por meio do grupo focal, coletamos evidências sobre a importância do *Framework* para Avaliação da Saúde de Ecossistemas de Dados, uma vez que foi possível obter bons resultados e sugestões de melhorias relevantes para o FASED. Após essa avaliação, o modelo foi refinado e algumas das melhorias propostas pelos participantes foram incorporadas.

Por fim, como parte da pesquisa sobre Ecossistemas de Dados foram publicados os artigos:

- Santos, H.D.; Oliveira, I.S.; Lima, G.F.B.; Silva, K.M.; Cruz, R.V.; 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.
- Lima, G.F.B.; Oliveira, M.I.S.; Lóscio, B.F. (2018) **Avaliação da Saúde de Ecossistemas de Dados**. Workshop on Thesis and Dissertations in Databases - SBBDD.
- Oliveira, M.I.S.; Lima, G.F.B.; Lóscio, B.F. (2019) **Investigations into data ecosystems: a systematic mapping study**. Knowledge and Information Systems p. 1-42.

7.2 LIMITAÇÕES

Durante o estudo algumas limitações puderam ser observadas, mesmo com o esforço para mitigá-las por parte dos pesquisadores. Primeiramente, não foi fácil encontrar pesquisadores e profissionais para participar da avaliação do *framework*. A maioria dos participantes tinham alguma familiaridade com Ecossistemas de Dados, o que pode causar um certo viés na formação da opinião de cada um deles. Além disso, conciliar as agendas dos participantes foi algo muito complicado e, conseqüentemente, não foi possível criar uma grande amostra representativa da população.

Em relação ao método de pesquisa, as limitações são típicas de estudos empíricos, particularmente na generalização dos resultados. A extração e análise dos dados pode ter sido influenciada pelas opiniões pessoais do pesquisador que executou o processo. Para mitigar essa ameaça, dedicamos um tempo para executar vários ciclos de refinamento para construir o *framework* proposto. Outra ação de mitigação consistiu em consultar frequentemente outros pesquisadores do grupo de pesquisa ALADIN para abordar e resolver os conflitos apresentados no decorrer da pesquisa.

Com relação ao *framework*, mais aplicações práticas devem ser realizadas em diferentes domínios e, principalmente, em cenários reais, a fim de verificar a completude e adequação em diferentes contextos.

7.3 TRABALHOS FUTUROS

Como trabalhos futuros identificamos as seguintes oportunidades:

- **Replicar a avaliação:** a aplicação de outro grupo focal com outros participantes permitirá reunir novas evidências. Essas novas evidências poderão ser usadas para realizar melhorias no *framework*, quanto para o seu amadurecimento.
- **Automação da avaliação da saúde:** um trabalho relevante é o desenvolvimento de uma ferramenta que auxilie na coleta, cálculo e análise dos resultados, onde cada métrica possui um peso de acordo com o seu impacto na saúde do ED. A coleta pode ser automática, se houver APIs que forneçam as informações necessárias, e manual para as métricas em que a coleta não pode ser automática. Essa aplicação *Web* conterá também um *dashboard* para visualizar o panorama do estado da saúde ao longo do tempo, fornecendo informações sobre os indicadores que estão com deficiências, e especificando quais características e métricas precisam de mais atenção.
- **Aplicar o FASED em cenários reais:** como o FASED foi criado com o propósito de ser genérico, um trabalho futuro seria realizar a instanciação do FASED para outros domínios de dados (*e.g.* Dados Abertos, Dados Privados, Dados Científicos) e aplicá-lo nesses cenários reais onde podem aparecer situações específicas que não foram pensadas durante a sua construção.
- **Inclusão de métricas qualitativas:** as métricas encontradas na literatura e as propostas pelo FASED são em sua maioria quantitativas, mas métricas qualitativas podem ser incluídas para complementar a avaliação das características propostas pelo FASED, e também para avaliar novos aspectos, como a força entre os relacionamentos.
- **Inclusão de indicador financeiro:** durante a avaliação do *framework* foi apontada a necessidade de incluir um novo indicador que inclua somente métricas relacionadas ao aspecto financeiro do ED. Apesar dos indicadores existentes, Robustez e Sustentabilidade, já tratarem desses aspectos, como trabalho futuro pode ser feito um estudo mais aprofundado para identificar quais aspectos financeiros não são cobertos pelo FASED.

REFERÊNCIAS

- ALAN, R. H. V.; MARCH, S. T.; PARK, J.; RAM, S. Design science in information systems research. *MIS quarterly*, Springer, vol. 28, n. 1, p. 75–105, 2004.
- ATTARD, J.; ORLANDI, F.; AUER, S. Data value networks: Enabling a new data ecosystem. *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, p. 453–456, 2016.
- BARBOSA, L.; PHAM, K.; SILVA, C.; VIEIRA, M. R.; FREIRE, J. Structured open urban data: understanding the landscape. *Big data*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 2, n. 3, p. 144–154, 2014.
- BARRY, M.-L.; STEYN, H.; BRENT, A. Determining the most important factors for sustainable energy technology selection in africa: Application of the focus group technique. *PICMET 08 - 2008 Portland International Conference on Management of Engineering Technology*, p. 181–187, July 2008. ISSN 2159-5100.
- BEDOYA, Ó. H. F.; AMELLER, D.; COSTA, D. C.; GUTIÉRREZ, J. F. Queso v2.0 a quality model for open source software ecosystems: List of measures. 2016.
- BLACKBURN, R.; STOKES, D. Breaking down the barriers: Using focus groups to research small and medium-sized enterprises. *International Small Business Journal*, v. 19, n. 1, p. 44–67, 2000. Disponível em: <<https://doi.org/10.1177/0266242600191003>>.
- BOTELHO, R. G.; OLIVEIRA, C. d. C. de. Literaturas branca e cinzenta: uma revisão conceitual. *Ciência da Informação*, v. 44, n. 3, 2015.
- BOURNE, P. E.; LORSCH, J. R.; GREEN, E. D. Perspective: Sustaining the big-data ecosystem. *Nature*, v. 527, n. 7576, p. S16, 2015.
- CARVALHO, I.; CAMPOS, F.; BRAGA, R.; DAVID, J. M. N.; STROELLE, V.; ARAÚJO, M. A. Heal me-an architecture for health software ecosystem evaluation. *2017 IEEE/ACM Joint 5th International Workshop on Software Engineering for Systems-of-Systems and 11th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (JSOS)*, p. 59–65, 2017.
- CARVALHO, I.; VEIGA, W.; CAMPOS, F.; BRAGA, R.; STRÖELE, V. Qualidade de um ecossistema de e-learning: Indicadores de saúde. *Anais do XIII Simpósio Brasileiro de Sistemas de Informação*, p. 230–237, 2017.
- CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. *Mobile networks and applications*, Springer, v. 19, n. 2, p. 171–209, 2014.
- CHIARA, I. G. D. Grupo de foco. *Métodos qualitativos de pesquisa em Ciência da Informação*, Polis, São Paulo, Brasil, p. 101–117, 2005.
- CHRISTENSEN, H. B.; HANSEN, K. M.; KYNG, M.; MANIKAS, K. Analysis and design of software ecosystem architectures—towards the 4s telemedicine ecosystem. *Information and Software Technology*, Elsevier, v. 56, n. 11, p. 1476–1492, 2014.

- CHUN, S.; SHULMAN, S.; SANDOVAL, R.; HOVY, E. Government 2.0: Making connections between citizens, data and government. *Information Polity*, IOS Press, v. 15, n. 1, 2, p. 1–9, 2010.
- COSTANZA, R. Toward an operational definition of ecosystem health. *Ecosystem health: New goals for environmental management*, Island Press Washington, DC, p. 239–256, 1992. *Ecosystem health: New goals for environmental management*. 239–256.
- CRESWELL, J. W. Mapping the developing landscape of mixed methods research. *SAGE handbook of mixed methods in social behavioral research 2*, p. 45–68, 2010.
- DAVIES, T. Open data: infrastructures and ecosystems. *Open Data Research*, p. 1–6, 2011.
- DAWES, S. S.; VIDIASOVA, L.; PARKHIMOVICH, O. Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, Elsevier, v. 33, n. 1, p. 15–27, 2016.
- DHUNGANA, D.; GROHER, I.; SCHLUDERMANN, E.; BIFFL, S. Software ecosystems vs. natural ecosystems: learning from the ingenious mind of nature. *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume*, p. 96–102, 2010.
- DING, L.; LEBO, T.; ERICKSON, J. S.; DIFRANZO, D.; WILLIAMS, G. T.; LI, X.; MICHAELIS, J.; GRAVES, A.; ZHENG, J. G.; SHANGGUAN, Z. et al. Twc logd: A portal for linked open government data ecosystems. *Web Semantics: Science, Services and Agents on the World Wide Web*, Elsevier, vol. 9, n. 3, p. 325–333, 2011.
- DONKER, F. W.; LOENEN, B. van. How to assess the success of the open data ecosystem? *International Journal of Digital Earth*, Taylor & Francis, vol. 10, n. 3, p. 284–306, 2017.
- EASTERBROOK, S.; SINGER, J.; STOREY, M.-A.; DAMIAN, D. Selecting empirical methods for software engineering research. *Guide to advanced empirical software engineering*, Springer, p. 285–311, 2008.
- FRANCO-BEDOYA, O.; AMELLER, D.; COSTAL, D.; FRANCH, X. Queso a quality model for open source software ecosystems. *International Conference on Software Engineering and Applications*, p. 209–221, 2014.
- GAMA, K.; LÓSCIO, B. F. Towards ecosystems based on open data as a service. *International Conference on Enterprise Information Systems*, p. 659–664, 2014.
- GEERTS, G. L. A design science research methodology and its application to accounting information systems research. *International Journal of Accounting Information Systems*, Elsevier, v. 12, n. 2, p. 142–151, 2011.
- HA, S.; LEE, S.; LEE, K. Standardization requirements analysis on big data in public sector based on potential business models. *International Journal of Software Engineering and Its Applications*, v. 8, n. 11, p. 165–172, 2014.
- HANSSSEN, G. K.; DYBÅ, T. Theoretical foundations of software ecosystems. *IWSECO@ICSOB*, p. 6–17, 2012.

- HARRISON, T. M.; PARDO, T. A.; COOK, M. Creating open government ecosystems: A research and development agenda. *Future Internet*, Molecular Diversity Preservation International, v. 4, n. 4, p. 900–928, 2012.
- HARTIGH, E. den; TOL, M.; VISSCHER, W. The health measurement of a business ecosystem. *European Network on Chaos and Complexity Research and Management Practice Meeting*, p. 1–39, 2006.
- HEIMSTÄDT, M.; SAUNDERSON, F.; HEATH, T. Conceptualizing open data ecosystems: A timeline analysis of open data development in the uk. *Conference for E-Democracy and Open Government*, p. 245, 2014.
- HEVNER, A.; CHATTERJEE, S. Design science research in information systems. *Design research in information systems*, Springer, p. 9–22, 2010.
- HEVNER, A. R. A three cycle view of design science research. *Scandinavian journal of information systems*, v. 19, n. 2, p. 4, 2007.
- IANSTITI, M.; LEVIEN, R. Keystones and dominators: Framing the operational dynamics of business ecosystems. *The Operational Dynamics of Business Ecosystems*, 2002.
- IANSTITI, M.; LEVIEN, R. The keystone advantage: What the new dynamics of business ecosystems mean for strategy, innovation, and sustainability. *Harvard Business School Press*, 2004.
- IANSTITI, M.; LEVIEN, R. Strategy as ecology. *Harvard business review*, vol. 82, n. 3, p. 68–81, 2004.
- III, F. S. C.; TORN, M. S.; TATENO, M. Principles of ecosystem sustainability. *American Naturalist*, JSTOR, p. 1016–1037, 1996.
- IMMONEN, A.; PALVIAINEN, M.; OVASKA, E. Requirements of an open data based business ecosystem. *IEEE access*, IEEE, v. 2, p. 88–103, 2014.
- JABAREEN, Y. Building a conceptual framework: philosophy, definitions, and procedure. *International journal of qualitative methods*, SAGE Publications Sage CA: Los Angeles, CA, v. 8, n. 4, p. 49–62, 2009.
- JANSEN, S. Measuring the health of open source software ecosystems: Beyond the scope of project health. *Information and Software Technology*, Elsevier, v. 56, n. 11, p. 1508–1519, 2014.
- KEELE, S. et al. Guidelines for performing systematic literature reviews in software engineering. *Technical report, Ver. 2.3 EBSE Technical Report. EBSE*, 2007.
- KONTIO, J.; LEHTOLA, L.; BRAGGE, J. Using the focus group method in software engineering: obtaining practitioner and user experiences. *International Symposium on Empirical Software Engineering*, p. 271–280, 2004.
- KÖSTER, V.; SUÁREZ, G. Open data for development: experience of uruguay. *International Conference on Theory and Practice of Electronic Governance*, p. 207–210, 2016.

- KOZNOV, D.; ANDREEVA, O.; NIKULA, U.; MAGLYAS, A.; MUROMTSEV, D.; RADCHENKO, I. Open government data in russian federation. *arXiv preprint arXiv:1612.05164*, 2016.
- LEE, D. Building an open data ecosystem: an irish experience. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, p. 351–360, 2014.
- LINDMAN, J.; KINNARI, T.; ROSSI, M. Business roles in the emerging open-data ecosystem. *IEEE Software*, IEEE, v. 33, n. 5, p. 54–59, 2015.
- LUNDELL, B.; FORSSTEN, B.; GAMALIELSSON, J.; GUSTAVSSON, H.; KARLSSON, R.; LENNERHOLT, C.; LINGS, B.; MATTSSON, A.; OLSSON, E. Exploring health within oss ecosystems. *First international workshop on building sustainable open source communities*, p. 1–5, 2009.
- MAGALHAES, G.; ROSEIRA, C.; MANLEY, L. Business models for open government data. *International Conference on Theory and Practice of Electronic Governance*, p. 365–370, 2014.
- MANIKAS, K.; HANSEN, K. M. Reviewing the health of software ecosystems—a conceptual framework proposal. *International Workshop on Software Ecosystems*, p. 33–44, 2013.
- MANIKAS, K.; HANSEN, K. M. Software ecosystems—a systematic literature review. *Journal of Systems and Software*, Elsevier, v. 86, n. 5, p. 1294–1306, 2013.
- MARCONI, M. d. A.; LAKATOS, E. M. *Fundamentos de metodologia científica*. [S.l.]: Atlas, 2010.
- MERCADO-LARA, E.; GIL-GARCIA, J. R. Open government and data intermediaries: the case of aiddata. *International Conference on Digital Government Research*, p. 335–336, 2014.
- MOISO, C.; MINERVA, R. Towards a user-centric personal data ecosystem the role of the bank of individuals’ data. *International Conference on Intelligence in Next Generation Networks*, p. 202–209, 2012.
- NOY, N. F.; MCGUINNESS, D. L. et al. Ontology development 101: A guide to creating your first ontology. *Stanford University*, 2001.
- NUSEIBEH, B. Weaving together requirements and architectures. *Computer*, IEEE, v. 34, n. 3, p. 115–119, 2001.
- OLIVEIRA, M. I. S.; LIMA, G. d. F. A. B.; LÓSCIO, B. F. Investigations into data ecosystems: A systematic mapping study. *Knowledge and Information Systems*, 2019.
- OLIVEIRA, M. I. S.; LÓSCIO, B. F. What is a data ecosystem? *Digital Government Research*, 2018.
- OLIVEIRA, M. I. S.; OLIVEIRA, H. R. de; OLIVEIRA, L. A.; LÓSCIO, B. F. Open government data portals analysis: the brazilian case. *International Digital Government Research Conference on Digital Government Research*, p. 415–424, 2016.

- OLIVEIRA, M. I. S.; OLIVEIRA, L. E. R.; BATISTA, M. G. R.; LÓSCIO, B. F. Towards a meta-model for data ecosystems. *International Conference on Digital Government Research: Governance in the Data Age*, p. 72, 2018.
- Oxford Dictionary. *Definition of Framework*. [S.l.]: Oxford Dictionary. Available: www.oxforddictionaries.com, 2019.
- POLLOCK, R. Building the (open) data ecosystem. *Open knowledge foundation Blog*, v. 31, 2011.
- PUSCHMANN, C.; BURGESS, J. The politics of twitter data. HIIG Discussion Paper Series, 2013.
- RAPPORT, D. J.; COSTANZA, R.; MCMICHAEL, A. Assessing ecosystem health. *Trends in ecology & evolution*, Elsevier, v. 13, n. 10, p. 397–402, 1998.
- ROBERT, K. Y. Case study research: design and methods. *Sage Publications*, 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.
- SCHAEFFER, D. J.; HERRICKS, E. E.; KERSTER, H. W. Ecosystem health: I. measuring ecosystem health. *Environmental Management*, Springer, vol. 12, n. 4, p. 445–455, 1988.
- SCHALKWYK, F. V.; WILLMERS, M.; MCNAUGHTON, M. Viscous open data: The roles of intermediaries in an open data ecosystem. *Information Technology for Development*, Taylor & Francis, v. 22, n. sup1, p. 68–83, 2016.
- SHIN, D.-H. Demystifying big data: Anatomy of big data developmental process. *Telecommunications Policy*, Elsevier, v. 40, n. 9, p. 837–854, 2016.
- SHIN, D.-H.; CHOI, M. J. Ecological views of big data: Perspectives and issues. *Telematics and Informatics*, Elsevier, v. 32, n. 2, p. 311–320, 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.
- SMITH, G.; OFE, H. A.; SANDBERG, J. Digital service innovation from open data: Exploring the value proposition of an open data marketplace. *International Conference on System Sciences*, p. 1277–1286, 2016.
- SOTO, M.; CIOLKOWSKI, M. The qualoss open source assessment model measuring the performance of open source communities. *International Symposium on Empirical Software Engineering and Measurement*, p. 498–501, 2009.
- STRONG, D. M.; LEE, Y. W.; WANG, R. Y. Data quality in context. *Communications of the ACM*, ACM, v. 40, n. 5, p. 103–110, 1997.
- TEECE, D. J. Business models, business strategy and innovation. *Long range planning*, Elsevier, v. 43, n. 2-3, p. 172–194, 2010.

- UBALDI, B. Open government data. *OECD*, 2013.
- VENABLE, J.; PRIES-HEJE, J.; BASKERVILLE, R. A comprehensive framework for evaluation in design science research. *International Conference on Design Science Research in Information Systems*, p. 423–438, 2012.
- 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.
- ZELETI, F. A.; OJO, A. Capability matrix for open data. In: SPRINGER. *Working Conference on Virtual Enterprises*. [S.l.], 2014. p. 498–509.
- ZELETI, F. A.; OJO, A. Critical factors for dynamic capabilities in open government data enabled organizations. *International Digital Government Research Conference on Digital Government Research*, p. 86–96, 2016.
- ZELETI, F. A.; OJO, A. Open data value capability architecture. *Information Systems Frontiers*, Springer, v. 19, n. 2, p. 337–360, 2017.
- ZOTT, C.; AMIT, R. Business model design: an activity system perspective. *Long range planning*, Elsevier, v. 43, n. 2-3, p. 216–226, 2010.
- ZUBCOFF, J.; GREGORI, L. V.; LÓPEZ, J. N. M.; PÉREZ, F. M.; GARRIGÓS, I.; FUSTER-GUILLÓ, A.; ALCOVER, J. V. C. et al. The university as an open data ecosystem. WIT Press, 2016.
- ZUIDERWIJK, A.; JANSSEN, M.; CHOENNI, S.; MEIJER, R.; ALIBAKS, R. S. Socio-technical impediments of open data. *Electronic Journal of e-Government*, v. 10, n. 2, 2012.
- ZUIDERWIJK, A.; JANSSEN, M.; DAVIS, C. Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, IOS Press, v. 19, n. 1, 2, p. 17–33, 2014.
- ZUIDERWIJK, A.; JANSSEN, M.; KAA, G. van de; POULIS, K. The wicked problem of commercial value creation in open data ecosystems: Policy guidelines for governments. *Information Polity*, IOS Press, vol. 21, n. 3, p. 223–236, 2016.
- ZUIDERWIJK, A.; JANSSEN, M.; POULIS, K.; KAA, G. van de. Open data for competitive advantage: insights from open data use by companies. *International Conference on Digital Government Research*, p. 79–88, 2015.

APÊNDICE A – FORMULÁRIO DE AVALIAÇÃO DO FASED

Meu nome é Glória de Fátima Andrade Barros Lima. 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 Framework para Avaliação da Saúde de Ecossistemas de Dados (FASED). Em particular, o framework proposto é composto por quatro indicadores de saúde, cada um contendo um conjunto de características e cada característica, por sua vez, é constituída de uma ou mais métricas. O framework tem o objetivo de avaliar a saúde de ecossistemas de dados por meio de métricas quantitativas e binárias que medem o estado atual da saúde do ecossistema e a capacidade de permanecer saudável ao longo do tempo.

Suas respostas são anônimas, ou seja, o resultado da sua participação será identificado por um código e poderão ser utilizados fragmentos da sua fala e/ou escrita no trabalho.

Você não precisa fornecer suas informações pessoais, além do seu e-mail, se quiser receber os resultados dessa pesquisa.

FASE 1 - Caracterização do Participante

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. Nome:

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

() Graduação

() Pós-graduação - Especialização

() Pós-graduação - Mestrado Acadêmico/Profissional

() Pós-graduação - Doutorado

() Pós-doutorado

3. Qual o seu cargo/função atualmente?

4. Você desenvolve suas atividades em qual meio?

- Acadêmico
- Profissional
- Ambos

5. Como você classificaria o seu grau de conhecimento sobre Ecossistema de Dados de uma forma geral? Numa escala de 1 a 10, na qual 1 significa “Nenhum” e 10 significa “Avançado”.

6. Você já participou ou participa de algum Ecossistema de Dados?

- Sim
- Não

7. Se sim, como você caracteriza sua participação?

- Atuo como consumidor de dados
- Atuo como publicador de dados
- Atuo no desenvolvimento de soluções e serviços
- Atuo na gestão do ecossistema
- Atuo como consultor/auditor
- Atuo na formulação de políticas/ações para Ecossistemas de Dados

8. Se sim, por quanto tempo atuou ou tem atuado em Ecossistemas de Dados?

- Menos de 2 anos
- De 2 a menos de 3 anos
- De 3 a menos de 4 anos
- De 4 a menos de 5 anos
- Mais de 5 anos

FASE 2 - Avaliação dos Elementos do FASED

O Framework para Avaliação da Saúde de Ecossistemas de Dados denominado FASED é composto por 4 indicadores de saúde, cada indicador contém entre 2 e 5 características e cada característica é constituída de uma ou mais métricas. O framework tem o objetivo de avaliar a saúde de Ecossistemas de Dados por meio de métricas quantitativas e binárias que medem o estado atual da saúde do ecossistema e a capacidade do ecossistema em permanecer saudável ao longo do tempo.

Indicadores

Os indicadores correspondem à interpretação dos resultados de um conjunto de métricas aplicadas que influenciam nas tomadas de decisão do ecossistema. A presença do conjunto de indicadores indica um panorama favorável da saúde do ecossistema analisado.

1. O conjunto de indicadores é adequado para a avaliação da saúde de Ecossistemas de Dados?

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

- Sim, o conjunto de indicadores é adequado para avaliar a saúde de Ecossistemas de Dados
- Não, um ou mais indicadores precisam ser incluídos
- Não, um ou mais indicadores precisam ser excluídos

Quais são as suas sugestões em caso de necessidade de inclusão ou exclusão de algum indicador?

2. A descrição dos indicadores propostos no FASED é clara?

A ideia aqui é identificar se os indicadores são fáceis de serem compreendidos. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- Sim, a descrição dos indicadores é clara
- Não, um ou mais indicadores precisam ser alterados

Quais são as suas sugestões para melhorar os indicadores?

Características

As características correspondem a um aspecto abstrato que compõem o panorama do indicador ao qual ela pertence. Cada indicador é composto por 2 a 5 características. Cada característica possui um conjunto de métricas relacionadas que medem aspectos concretos do ecossistema.

Características - Indicador: Produtividade

1. O conjunto de características do indicador de Produtividade é adequado para a avaliação da saúde de Ecossistemas de Dados?

A ideia aqui é identificar se o conjunto de características é suficiente e cobrem os aspectos abstratos relacionados com o indicador. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- () Sim, o conjunto de características é adequado para avaliar a saúde de Ecossistemas de Dados
- () Não, uma ou mais características precisam ser incluídas
- () Não, uma ou mais características precisam ser excluídas

Quais são as suas sugestões em caso de necessidade de inclusão ou exclusão de alguma característica?

2. A descrição das características do indicador de Produtividade é clara?

A ideia aqui é identificar se as características são fáceis de serem compreendidas. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- () Sim, a descrição das características é clara
- () Não, uma ou mais características precisam ser alteradas

Quais são as suas sugestões em caso de alteração?

Características - Indicador: Robustez

1. O conjunto de características do indicador de Robustez é adequado para a avaliação da saúde de Ecossistemas de Dados?

A ideia aqui é identificar se o conjunto de características é suficiente e cobrem os aspectos abstratos relacionados com o indicador. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- () Sim, o conjunto de características é adequado para avaliar a saúde de Ecossistemas de Dados
- () Não, uma ou mais características precisam ser incluídas
- () Não, uma ou mais características precisam ser excluídas

Quais são as suas sugestões em caso de necessidade de inclusão ou exclusão de alguma característica?

2. A descrição das características do indicador de Robustez é clara?

A ideia aqui é identificar se as características são fáceis de serem compreendidas. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- () Sim, a descrição das características é clara
() Não, uma ou mais características precisam ser alteradas

Quais são as suas sugestões em caso de falta de clareza?

Características - Indicador: Criação de Nicho**1. O conjunto de características do indicador de Criação de Nicho é adequado para a avaliação da saúde de Ecossistemas de Dados?**

A ideia aqui é identificar se o conjunto de características é suficiente e cobrem os aspectos abstratos relacionados com o indicador. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- () Sim, o conjunto de características é adequado para avaliar a saúde de Ecossistemas de Dados
() Não, uma ou mais características precisam ser incluídas
() Não, uma ou mais características precisam ser excluídas

Quais são as suas sugestões em caso de necessidade de inclusão ou exclusão de alguma característica?

2. A descrição das características do indicador de Criação de Nicho é clara?

A ideia aqui é identificar se as características são fáceis de serem compreendidas. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- () Sim, a descrição das características é clara
() Não, uma ou mais características precisam ser alteradas

Quais são as suas sugestões em caso de falta de clareza?

Características - Indicador: Sustentabilidade**1. O conjunto de características do indicador de Sustentabilidade é adequado para a avaliação da saúde de Ecossistemas de Dados?**

A ideia aqui é identificar se o conjunto de características é suficiente e cobrem os aspectos abstratos relacionados com o indicador. Também gostaríamos de identificar lacunas, erros e/ou possibilidades de melhoria.

- () Sim, o conjunto de características é adequado para avaliar a saúde de Ecossistemas de Dados
- () Não, uma ou mais características precisam ser incluídas
- () Não, uma ou mais características precisam ser excluídas

Quais são as suas sugestões em caso de necessidade de inclusão ou exclusão de alguma característica?

2. A descrição das características do indicador de Sustentabilidade é clara?

A ideia aqui é identificar se as características são fáceis de serem compreendidas. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- () Sim, a descrição das características é clara
- () Não, uma ou mais características precisam ser alteradas

Quais são as suas sugestões em caso de falta de clareza?

Métricas

As métricas nos ajudam a medir um aspecto concreto do Ecossistema de Dados e estão relacionadas a atributos do Ecossistema (atores, recursos, papéis e relações). Os resultados obtidos por meio da aplicação das métricas auxiliam no gerenciamento de riscos, detecção de problemas e no panorama da saúde do ecossistema. As métricas propostas no FASED foram extraídas de frameworks que avaliam a saúde de ecossistemas de software, de modelos de qualidade de software, de trabalhos de qualidade de dados e outras foram de autoria da própria autora para suprir necessidades específicas do contexto de Ecossistemas de Dados.

Métricas - Indicador: Produtividade

1. O conjunto de métricas é adequado para medir a Produtividade?

A ideia aqui é identificar se o conjunto de métricas é suficiente e avaliam a produtividade do ecossistema em estudo. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- () Sim, o conjunto de métricas é adequado

() Não, uma ou mais métricas precisam ser incluídas

() Não, uma ou mais métricas precisam ser excluídas

Quais são as suas sugestões em caso de necessidade de inclusão ou exclusão das métricas?

2. O detalhamento das métricas é de fácil entendimento e facilita sua aplicabilidade?

A ideia aqui é identificar se o detalhamento das métricas facilita seu entendimento e auxilia na aplicação delas. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

() Sim, as métricas são de fácil entendimento e aplicabilidade

() Não, uma ou mais informações das métricas precisam ser alteradas

Quais são as suas sugestões para melhorar a compreensão e a aplicação das métricas?

Métricas - Indicador: Robustez

1. O conjunto de métricas é adequado para medir a Robustez?

A ideia aqui é identificar se o conjunto de métricas é suficiente e avaliam a robustez do ecossistema em estudo. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

() Sim, o conjunto de métricas é adequado

() Não, uma ou mais métricas precisam ser incluídas

() Não, uma ou mais métricas precisam ser excluídas

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

2. O detalhamento das métricas é de fácil entendimento e facilita sua aplicabilidade?

A ideia aqui é identificar se o detalhamento das métricas facilita seu entendimento e auxilia na aplicação delas. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

() Sim, as métricas são de fácil entendimento e aplicabilidade

() Não, uma ou mais informações das métricas precisam ser alteradas

Quais são as suas sugestões para melhorar a compreensão e a aplicação das métricas?

Métricas - Indicador: Criação de Nicho

1. O conjunto de métricas é adequado para medir a Criação de Nicho?

A ideia aqui é identificar se o conjunto de métricas é suficiente e avaliam a criação de nicho do ecossistema em estudo. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- () Sim, o conjunto de métricas é adequado
- () Não, uma ou mais métricas precisam ser incluídas
- () Não, uma ou mais métricas precisam ser excluídas

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

2. O detalhamento das métricas é de fácil entendimento e facilita sua aplicabilidade?

A ideia aqui é identificar se o detalhamento das métricas facilita seu entendimento e auxilia na aplicação delas. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- () Sim, as métricas são de fácil entendimento e aplicabilidade
- () Não, uma ou mais informações das métricas precisam ser alteradas

Quais são as suas sugestões para melhorar a compreensão e a aplicação das métricas?

Métricas - Indicador: Sustentabilidade

1. O conjunto de métricas é adequado para medir a Sustentabilidade?

A ideia aqui é identificar se o conjunto de métricas é suficiente e avaliam a sustentabilidade do ecossistema em estudo. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

- () Sim, o conjunto de métricas é adequado
- () Não, um ou mais métricas precisam ser incluídas
- () Não, um ou mais métricas precisam ser excluídas

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

2. O detalhamento das métricas é de fácil entendimento e facilita sua aplicabilidade?

A ideia aqui é identificar se o detalhamento das métricas facilita seu entendimento e auxilia na aplicação delas. Também gostaríamos de identificar erros e/ou possibilidades de melhoria.

() Sim, as métricas são de fácil entendimento e aplicabilidade

() Não, uma ou mais informações das métricas precisam ser alteradas

Quais são as suas sugestões para melhorar a compreensão e a aplicação das métricas?

FASE 3 - Perguntas Gerais

1. O FASED é relevante para avaliar a saúde de Ecossistemas de Dados.

A ideia aqui é identificar se, para você, o FASED é aplicável, útil e importante para avaliar a saúde de Ecossistemas de Dados. Numa escala de 1 a 5, na qual 1 significa “Discordo Fortemente” e 5 significa “Concordo Fortemente”.

Comente aqui a sua opinião sobre a relevância do FASED.

2. O FASED é genérico o suficiente para avaliar a saúde de Ecossistemas de Dados de qualquer domínio.

A ideia aqui é identificar se, para você, o FASED é genérico o suficiente para ser aplicado em qualquer domínio de Ecossistemas de Dados. Numa escala de 1 a 5, na qual 1 significa “Discordo Fortemente” e 5 significa “Concordo Fortemente”.

Comente aqui a sua opinião sobre o quão genérico é o FASED.

3. O FASED é adaptável para qualquer domínio de Ecossistemas de Dados.

A ideia aqui é identificar se, para você, o FASED é genérico o suficiente para ser aplicado em qualquer domínio de Ecossistemas de Dados. Numa escala de 1 a 5, na qual

1 significa “Discordo Fortemente” e 5 significa “Concordo Fortemente”.

Comente aqui a sua opinião sobre o quão adaptável é o FASED.

4. A hierarquia de elementos do FASED é coerente.

A ideia aqui é identificar se, para você, a hierarquia do FASED, a qual é composta por indicadores, características e métricas é coerente. Numa escala de 1 a 5, na qual 1 significa “Discordo Fortemente” e 5 significa “Concordo Fortemente”.

Comente aqui a sua opinião sobre a coerência da hierarquia do FASED.
