

Em Direção a uma Abordagem para Avaliação da Saúde de Ecossistemas de Dados

Glória de Fátima B. Lima¹, Marcelo Iury S. Oliveira^{1,2}, Bernadette Farias Lóscio¹

¹Centro de Informática -- Universidade Federal de Pernambuco (UFPE)

²Unidade Acadêmica de Serra Talhada – Universidade Federal Rural de Pernambuco

{gfabl,miso,bfl}@cin.ufpe.br, marcelo.iury@ufrpe.br

Abstract. *The growing availability of data in digital media has contributed to the creation of a large number of data ecosystems. However, having successful Data Ecosystem is still a challenge. In order to prevent the failure of a Data Ecosystem and ensure its survival, evaluating its health becomes fundamental. In this paper, we propose an approach for Data Ecosystems health evaluation composed of a set of indicators and metrics, which assess the Data Ecosystem's current state and its ability to stay healthy over time.*

Resumo. *A crescente disponibilidade de dados em mídia digital tem contribuído para a criação de um grande número de ecossistemas de dados. No entanto, ter um ecossistema de dados bem-sucedido ainda é um desafio. Para prevenir a falha de um Ecossistema de Dados e garantir sua sobrevivência, avaliar sua saúde torna-se fundamental. Neste artigo, propomos uma abordagem para avaliação da saúde de Ecossistemas de Dados composta por um conjunto de indicadores e métricas que avaliam o estado atual do ecossistema de dados e sua capacidade de se manter saudável ao longo do tempo.*

1. Introdução

Ecossistemas de dados (EDs) são um novo ambiente composto por redes complexas de organizações e indivíduos que trocam e usam dados como recurso principal [Oliveira and Lóscio 2018]. Esses ecossistemas fornecem um ambiente para criar, gerenciar e sustentar iniciativas de compartilhamento de dados. 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.

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.

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 et al. 2019]. Os indicadores são avaliados 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) [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).

Nesse contexto, este trabalho propõe uma abordagem para avaliação da saúde de EDs baseada em um conjunto de indicadores, características e métricas que permitem avaliar o ED baseado nos 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.

Este trabalho está organizado da seguinte forma: a seção 2 apresenta os trabalhos relacionados, a seção 3 apresenta nossa proposta, a seção 4 detalha a avaliação da abordagem utilizando o método Focus Group e a seção 5 discute os próximos passos desta pesquisa. .

2. Trabalhos Relacionados

Até agora, as alternativas para avaliar a saúde dos ecossistemas 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 número de visualizações de página, downloads [Dawes et al. 2016]. Até onde sabemos, não existem estudos focados no uso de métricas mais elaboradas para avaliar a saúde de EDs [Oliveira et al. 2019]. No entanto, outros domínios do ecossistema possuem pesquisas mais maduras relacionadas ao tópico.

Há vasto material voltado a avaliação da saúde de Ecossistemas de Software e Ecossistemas de Negócio. O trabalho apresentado por [den Hartigh et al. 2006] propõe uma definição para a saúde dos Ecossistemas de Negócios e um conjunto de métricas que podem ser utilizadas pelos gestores para avaliar a saúde dos Ecossistemas de Negócios. Este trabalho tem como referência os construtos de saúde dos Ecossistemas Empresariais identificados por [Jansiti and Levien 2002], que são: produtividade, robustez e criação de nicho.

Para Ecossistemas de Software, o trabalho apresentado por [Jansen 2014] visa definir e analisar o funcionamento de Ecossistemas de Software Open Source, com base na produtividade, robustez e criação de nicho. Alternativamente, [Franco-Bedoya et al. 2014] propõe um modelo, denominado QuESo, composto por dois tipos de elementos: características de qualidade e métricas. As características de qualidade correspondem aos atributos de software relevantes para a avaliação, organizados em uma hierarquia composta no nível mais alto por 3 dimensões: a plataforma do Ecossistema de Software, a comunidade e sua rede de atores. Cada dimensão é dividida em características e estas, por sua vez, são divididas em subcaracterísticas. Outra proposta foi apresentada por [Carvalho et al. 2017] que propõe uma arquitetura de três camadas para avaliar a saúde dos ecossistemas de software, chamada HEAL ME. A arquitetura é composta pelas camadas de comunicação, serviço e descoberta de conhecimento.

Os estudos apresentados advogam sobre a importância de avaliar a saúde e o impacto dessa avaliação no desempenho e sustentabilidade do ecossistema. 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 dos EDs e atividades relacionadas, para possibilitar a avaliação da saúde do ecossistema em avaliação.

3. Abordagem Proposta

A abordagem proposta 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, a abordagem proposta contribuirá para: (i) auxiliar os atores tornarem seus EDs mais 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.

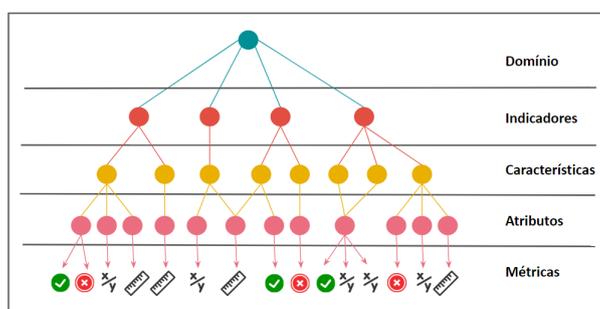


Figura 1. Abordagem de Avaliação de Saúde. Fonte: Autores

A abordagem de avaliação consiste na definição de um conjunto de indicadores, características, atributos e métricas que podem ser considerados pelos atores que vão executar a avaliação da saúde dos seus EDs. A utilização da abordagem descritiva permite a sua adaptação 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 EDs vão sendo mapeados.

A Figura 1 ilustra a abordagem proposta, 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 constitui o *framework*. Os conceitos de "indicador", "característica" e "métrica" são baseados na ISO/IEC 25000:2014 e no trabalho de [Soto and 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.

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 [Jansiti and Levien 2004] 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. 2017] [Jansen 2014] [Manikas and Hansen 2013] [den Hartigh et al. 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. 2017] como um novo indicador para avaliação da saúde de ecossistemas de software.

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. Uma alternativa é levantar as características a partir de um mapeamento/revisão sistemática de literatura. Por sua vez, os atributos podem ser os constructos principais de um Ecossistema de Dados. Os atributos podem ser selecionados a partir de trabalhos que buscam a proposição de meta-modelos para EDs ou que definem formalmente o que EDs, tais como [Oliveira et al. 2018, Oliveira and Lóscio 2018]. No nível final, as características e atributos são derivados em Métricas que são uma medida quantificável usada para rastrear e avaliar o status de uma característica específica. Enquanto uma métrica avalia uma característica, um indicador usará uma ou mais métricas para medir algum tópico. Um exemplo de documentação métrica é apresentado na Tabela 1.

Tabela 1. Métrica de produtividade para avaliar a saúde da DAECO

ID	Indicador	Métrica	Descrição	Medida e Fórmula	Interpretação	Unidade	Atributo	Referencia Bibliográfica
P1	Produtividade	Volume de dados produzidos	Volume de dados produzidos pelo ED e disponíveis para uso	$X = \sum_1^T N_T$ $N_T = \text{Volume de dados}$ $T = \text{Período de Tempo}$	$X \geq P$ $P = \text{Parâmetro}$ Quanto maior o volume de dados, melhor	Un - Unit	Resources	Author (2019)

4. Avaliação

Este artigo usou o método Grupo Focal para avaliar a viabilidade, coerência e adequação da abordagem. De acordo com [Kontio et al. 2004], um grupo focal é um técnica que envolve um grupo de participantes reunidos para discutir um determinado problema ou avaliar um determinado tópico. De forma a auxiliar na materialização da abordagem proposta, bem como melhor avaliá-la, o grupo focal utilizou como objeto de pesquisa o Framework FASED [Lima 2019] que implementa a abordagem proposta.

O grupo focal realizado nesta pesquisa foi composto por oito profissionais. 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.

Pedimos aos participantes que indicassem em que medida as seguintes afirmações são válidas: (i) O FASED é relevante para avaliar a saúde de Ecossistemas de Dados; (ii) O FASED é genérico o suficiente para avaliar a saúde de Ecossistemas de Dados de qualquer domínio; (iii) O FASED é adaptável para qualquer domínio de Ecossistemas de Dados; (iv) A hierarquia de elementos do FASED é coerente. Os participantes deviam indicar um valor de 1 a 5, onde o número 1 corresponde a “Discordo totalmente” e o número 5 corresponde a “Concordo totalmente”.

No geral, os participantes se mostraram satisfeitos com o *framework* proposto e relataram a importância dele para o contexto de avaliação da saúde de EDs. Com relação a sua relevância, sete dos participantes avaliaram como 5 e um participante avaliou com nota 4, tendo como resultado uma média geral de 4.875 nessa questão. Quanto a sua adaptabilidade, sete participantes avaliaram o FASED com 5, e um avaliou com nota 4, resultando em uma média geral de 4.875. E, por fim, quanto a generalidade, cinco participantes avaliaram o FASED com nota 5 e dois com nota 4, resultando em uma média geral de 4.75.

Os participantes que avaliaram a generalidade com nota 4 fizeram os seguintes comentários: “*Algumas métricas são muito 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.*” e outro participante deixou o seguinte comentário: “*A adaptação de algumas métricas e termos podem auxiliar numa maior generalidade*”.

5. Conclusão

Neste trabalho, propomos uma abordagem para Avaliação da Saúde de Ecossistemas de Dados baseada em na ISO/IEC 25000:2014 e de alguns trabalhos identificados durante o estudo ad-hoc. A abordagem proposta 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.

Como trabalho futuro, pretendemos realizar outro grupo focal com outros participantes, a fim de reunir novas evidências. Essas novas evidências podem ser utilizadas para fazer melhorias no *framework*, quanto ao seu amadurecimento.

Referências

Carvalho, I., Campos, F., Braga, R., David, J. M. N., Stroelle, V., and Araújo, M. A. (2017). 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)*, pages 59–65.

- Costanza, R. (1992). Toward an operational definition of ecosystem health. *Ecosystem health: New goals for environmental management*, pages 239–256. *Ecosystem health: New goals for environmental management*. 239–256.
- Dawes, S. S., Vidiiasova, L., and Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, 33(1):15–27.
- den Hartigh, E., Tol, M., and Visscher, W. (2006). The health measurement of a business ecosystem. *European Network on Chaos and Complexity Research and Management Practice Meeting*, pages 1–39.
- Dhungana, D., Groher, I., Schludermann, E., and Biffi, S. (2010). Software ecosystems vs. natural ecosystems: learning from the ingenious mind of nature. *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume*, pages 96–102.
- Franco-Bedoya, O., Ameller, D., Costal, D., and Franch, X. (2014). Queso a quality model for open source software ecosystems. *International Conference on Software Engineering and Applications*, pages 209–221.
- Iansiti, M. and Levien, R. (2002). Keystones and dominators: Framing the operational dynamics of business ecosystems. *The Operational Dynamics of Business Ecosystems*.
- Iansiti, M. and Levien, R. (2004). Strategy as ecology. *Harvard business review*, vol. 82(3):68–81.
- Jansen, S. (2014). Measuring the health of open source software ecosystems: Beyond the scope of project health. *Information and Software Technology*, 56(11):1508–1519.
- Kontio, J., Lehtola, L., and Bragge, J. (2004). Using the focus group method in software engineering: obtaining practitioner and user experiences. *International Symposium on Empirical Software Engineering*, pages 271–280.
- Lima, G. d. F. A. B. (2019). Um framework para avaliação da saúde de ecossistemas de dados. Master's thesis, Universidade Federal de Pernambuco, Centro de Informática, Curso de Pós-Graduação em Ciências da Computação, Recife.
- Manikas, K. and Hansen, K. M. (2013). Reviewing the health of software ecosystems—a conceptual framework proposal. *International Workshop on Software Ecosystems*, pages 33–44.
- Oliveira, M. I. S., Lima, G. d. F. A. B., and Lóscio, B. F. (2019). Investigations into data ecosystems: A systematic mapping study. *Knowledge and Information Systems*.
- Oliveira, M. I. S. and Lóscio, B. F. (2018). What is a data ecosystem? *Digital Government Research*.
- Oliveira, M. I. S., Oliveira, L. E. R., Batista, M. G. R., and Lóscio, B. F. (2018). Towards a meta-model for data ecosystems. *International Conference on Digital Government Research: Governance in the Data Age*, page 72.
- Soto, M. and Ciolkowski, M. (2009). The qualoss open source assessment model measuring the performance of open source communities. *International Symposium on Empirical Software Engineering and Measurement*, pages 498–501.