

## Avaliação da Saúde de Ecossistemas de Dados

**Aluna: Glória de Fátima Andrade Barros Lima**

E-mail: gfabl@cin.ufpe.br

**Orientador: Bernadette Farias Lóscio**

E-mail: bfl@cin.ufpe.br

**Co-orientador: Marcelo Iury S. Oliveira**

E-mail: miso@cin.ufpe.br

**Nível:** Mestrado

**Universidade Federal de Pernambuco - UFPE**

**Programa de Pós-graduação em Ciência da Computação – Centro de Informática**

**Mês e Ano de ingresso:** Março de 2017

**Mês e Ano previstos para defesa:** Março 2019

**Etapas Concluídas:** Créditos em disciplinas, Definição do tema e Levantamento do Referencial Bibliográfico

**Etapas Futuras:** Finalizar o levantamento e especificação dos critérios e métricas para avaliação da saúde de Ecossistemas de Dados, Formalizar a estratégia de avaliação de saúde em Ecossistemas de Dados, Avaliação da estratégia proposta.

***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 atualmente disponíveis, nem todos podem ser considerados saudáveis e sustentáveis, dificultando o compartilhamento e a troca 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 sustentabilidade ao longo do tempo. A saúde de um ED pode ser avaliada por meio de critérios, como: sustentabilidade, resiliência e geração de valor. Na literatura não foram encontrados trabalhos com este propósito. Neste trabalho, propomos uma estratégia para avaliação da saúde de Ecossistema de Dados consistindo de um conjunto de critérios e métricas que permitem uma avaliação contínua da saúde do ecossistema. Com o uso da estratégia proposta, espera-se obter um conjunto de 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.*

## 1. Introdução e Motivação

Nos últimos anos, estudos voltados para Ecossistemas de Dados (ED) vêm crescendo por causa da relevância que os dados apresentam nas tomadas de decisão em diferentes contextos, desde ambientes governamentais até ambientes corporativos. A disponibilização de dados por governos e entidades privadas tem crescido e facilitado o consumo, a troca e a produção de recursos que utilizam esses dados como meio para gerar valor e alavancar o crescimento. Um ED pode ser visto como “uma rede de 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 and Lóscio 2018].

Como a dimensão e a complexidade dos EDs são variáveis, a identificação de aspectos que impactam negativamente nesses ecossistemas se torna uma tarefa desafiadora. Sendo assim, como forma de identificar deficiências no funcionamento dos EDs, faz-se necessário ter indicadores que reflitam a saúde dos EDs. Neste trabalho, definimos como saúde do ED a sua capacidade de persistir e permanecer produtivo ao longo do tempo, tolerar a dinamicidade e variabilidade dos componentes e suprir as necessidades dos atores envolvidos.

Por se tratar de um tema recente, não existem trabalhos que avaliem especificamente a saúde dos EDs [Oliveira et al. 2017]. De maneira geral, a avaliação da saúde de ecossistemas é citada na literatura como uma forma de produzir indicadores operacionais sobre as atividades do ecossistema, os status dos elementos que o compõem, bem como gerar diretrizes sobre a governança do ecossistema [Oliveira et al. 2017] [Alves et al. 2017]. Entretanto, mais uma vez ressaltamos que a formalização de métricas específicas para avaliar a saúde dos EDs é um tópico ainda não explorado.

Nesse contexto, este trabalho propõe uma estratégia para avaliação da saúde de EDs, consistindo de um conjunto de dimensões, critérios e métricas que permitem avaliar o ED e, como consequência, geram um conjunto de indicadores sobre a saúde do ecossistema. Tendo em vista a ausência de trabalhos nesta área, usamos como inspiração propostas para avaliação da saúde de ecossistemas de negócios, ecossistemas de software e ecossistemas naturais.

Este trabalho está organizado como segue: a Seção 2 introduz os trabalhos relacionados, a Seção 3 apresenta nossa proposta, a Seção 4 detalha o método de pesquisa e a Seção 5 discute os próximos passos para a conclusão do trabalho.

## 2. Trabalhos Relacionados

[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. [Schaeffer et al. 1988] caracterizam um ecossistema em termos de estrutura e função, 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.

Na literatura de ecossistemas de negócios, o conceito de saúde é definido como a habilidade desses ecossistemas proverem oportunidades de crescimento duráveis para seus membros e para aqueles que dependem dele [Iansiti and Levien 2004a]. [Iansiti and Levien 2004b] definem três escalas, conhecidas como PRN, para representar a saúde de ecossistemas de negócios: a produtividade, a robustez e a criação de nicho ou inovação, as quais são avaliadas por meio de métricas que ponderam a habilidade de converter materiais brutos em novos produtos e de menor custo, de ser resiliente e de aumentar a diversidade de atores ao longo do tempo. Já [den Hartigh et al. 2006], inspirados pelos trabalhos iniciais da área, definem a saúde de ecossistemas de negócios como o bem-estar financeiro e a força da rede de conexões, e dividem a saúde em dois componentes: a saúde do parceiro e a saúde da rede. A saúde do parceiro avalia a saúde de cada ator individualmente e a saúde da rede de conexões é medida pela conectividade dos atores.

Na literatura de ecossistemas de software, a principal fonte de inspiração são os ecossistemas de negócios, pois ambos possuem parâmetros similares de avaliação de saúde. [Manikas and Hansen 2013] estabelece um *framework* para avaliação da saúde de ecossistemas de software definindo os atores, o software e o *orchestration* como os três principais componentes que afetam a saúde desses ecossistemas, eles consideram a saúde dos componentes de forma individual e coletiva. De maneira mais específica, [Wahyudin et al. 2007] definem três critérios que influenciam a saúde de projetos *open source*, que são a vivacidade da comunidade de desenvolvedores, a vivacidade da comunidade de usuários e a qualidade do produto, no qual cada critério é avaliado por métricas específicas.

Entretanto, nenhuma das soluções apresentadas se adéqua totalmente à realidade dos EDs por causa dos componentes e das relações existentes nesses ecossistemas. Por outro lado, o crescimento da quantidade de dados disponibilizados e consumidos direta ou indiretamente por atores diversos está alavancando a produção de novos recursos e revelando a necessidade de se ter ambientes saudáveis para lidar com essa nova dinâmica.

### 3. Ecossistemas de Dados

Como mencionado anteriormente, os EDs 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). No trabalho de [Oliveira and Lóscio 2018] foram identificados os principais componentes de um ED, os quais estão descritos a seguir. Estes elementos são a base para a escolha dos critérios e das métricas mais adequados para avaliação da saúde de um ED.

*Atores:* são entidades autônomas que exercem um ou mais papéis e são considerados elementos básicos do ED. Um conjunto de interesses motiva os atores e cada um tem expectativas diferentes. Os atores geralmente se comprometem com o ecossistema e precisam de incentivos para permanecer ativos no ecossistema.

*Recursos:* são os produtos, possessões ou capacidades úteis ou de valor, que são produzidos, providos, curados ou consumidos pelos atores. Os recursos podem ser trocados individualmente ou em conjunto por meio de transações.

*Papéis:* são funções exercidas por um ator no ED. Estão relacionadas com um

conjunto de deveres e atividades. Diversos papéis podem ser identificados nos ED, como consumidores, produtores e intermediários.

*Relacionamentos:* são as interações entre os atores, entre os recursos e entre os atores e os recursos dos EDs. Relacionamentos são geralmente baseados em interesses em comum ou são relacionados aos papéis que cada ator exerce no ecossistema.

Como cenário motivacional para exemplificar os componentes de um ED, suponha um ecossistema de dados de pesquisa acadêmico. Nesse cenário, os dados do contexto acadêmico são os protagonistas do ecossistema pois auxiliam os atores a produzir, trocar e consumir recursos. Os recursos oriundos de dados de pesquisa podem ser os próprios conjuntos de dados, os trabalhos acadêmicos ou os serviços e sistemas baseados nesses dados. Nesse ecossistema existem atores que lidam direta ou indiretamente com os recursos, alguns exemplos são os pesquisadores, entidades de pesquisa, financiadores, cidadãos, onde cada ator exerce um ou mais papéis, como os papéis de consumidor, produtor ou intermediário. Com isso, os relacionamentos entre os atores podem acontecer quando há interesses em comum (e.g. procura por um conjunto de dados específico) ou inerentes aos papéis (e.g. produção de soluções).

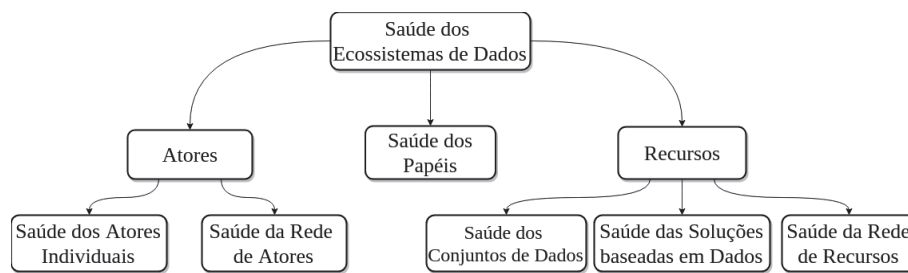
#### 4. Solução Proposta

Este trabalho propõe uma estratégia para Avaliação da Saúde de Ecossistemas de Dados, a qual objetiva a produção de indicadores que podem ser utilizados para avaliar aspectos qualitativos e quantitativos relacionados à produtividade, resiliência e geração de valor de um ED, por exemplo. A produção de indicadores é alcançada com a aplicação de métricas para medir aspectos e elementos em EDs reais.

Apesar do grande número de EDs atualmente disponíveis, nem todos são ecossistemas saudáveis e sustentáveis, dificultando assim o compartilhamento e a troca de dados pelos atores do ecossistema. Por isso, ter uma estratégia para avaliar a saúde dos EDs e identificar quais aspectos não estão com um bom desempenho é essencial para que o ecossistema consiga sobreviver e ser produtivo por mais tempo.

Para a definição de quais componentes representam os EDs nos baseamos no trabalho de [Oliveira and Lóscio 2018]. Complementando os componentes principais apresentados, nós consideramos que os componentes ator e recurso, precisavam ser divididos em componentes menores, pois possuem peculiaridades que afetam a saúde do ED de formas diferentes. A saúde dos atores pode ser dividida em saúde individual do ator e saúde da rede de atores, e similarmente, a saúde dos recursos pode ser dividida em saúde dos conjuntos de dados, saúde das soluções baseadas em dados e saúde da rede de recursos. Com base nos componentes identificados e no escopo de avaliação de saúde, nós decidimos tratar os componentes como dimensões, para que posteriormente pudéssemos atribuir critérios e métricas para avaliar essas dimensões. Sendo assim, na Figura 1 propomos o detalhamento das principais dimensões que influenciam na saúde geral dos EDs e explicamos a seguir cada uma.

*Saúde dos Atores Individuais:* Esta dimensão avalia como, em ecossistemas de negócios, a produtividade, a robustez e a geração de valor dos atores, ainda que individualmente, influenciam a saúde do ED. Os atores possuem características que podem ser medidas e assim ter a saúde avaliada, como: os atributos que os caracterizam, suas expectativas em relação ao ED, suas capacidades, os papéis que eles exercem e os recursos



**Figura 1. Dimensões para Avaliação de Saúde dos EDs**

produzidos, providos ou consumidos por eles. Na Figura 2 exemplificamos os critérios e aspectos que podem ser considerados no cálculo das métricas para avaliar a saúde da dimensão de atores individuais.

*Saúde da Rede de Atores:* A rede de atores e suas interações têm um papel importante na saúde dos EDs. Uma rede de atores compreende um conjunto de atores individuais que se relacionam, exercem papéis diferentes, seguem modelos de negócio para entregar valor e podem ter um ou mais atores-chave que coordenam suas transações. As interações existentes variam de acordo com o contexto (e.g. econômico, político, cultural e tecnológico), entretanto, elas possuem algumas características que podem ser avaliadas para definir a saúde da rede de atores, como os tipos de transações existentes, modelos de negócios, se há atores que coordenam a rede e qual o impacto disso.

	Critérios	Aspectos a serem Considerados
Atores Individuais	Produtividade	Ativo em quantos projetos; Transações que participa; Interações com outros EDs; Contexto inserido; Papéis que desempenha;
	Resiliência	Quanto tempo participa do ED; Dificuldades ao desempenhar seus papéis; Conflitos superados;
	Geração de Valor	Participação na entrega de recursos; Exerce papel de liderança; Utilização de modelos de negócio; Variedade nos projetos;

**Figura 2. Critérios para Avaliar a Saúde de Atores Individuais**

*Saúde dos Conjuntos de Dados:* Os conjuntos de dados são o agrupamento de dados relacionados ou não que têm a função principal de ser fonte de informação. A saúde dos conjuntos de dados pode ser avaliada utilizando métricas inspiradas nas melhores práticas dos dados na Web (*Data on the Web Best Practices*) [Lóscio et al. 2017], como a qualidade dos dados, formatos disponíveis e facilidade de acesso.

*Saúde das Soluções baseadas em Dados:* As soluções baseadas em dados são os recursos produzidos a partir do consumo dos conjuntos de dados, sejam elas aplicações, serviços, entre outros. A saúde das soluções baseadas em dados abrangem as métricas de saúde dos conjuntos de dados e também métricas que avaliam o processo de criação de ideias e desenvolvimento, entrega de valor e suas dificuldades.

*Saúde da Rede de Recursos:* Uma rede de recursos representa as relações de dependência entre esses recursos e a influência que eles exercem nas outras dimensões dos EDs. Por exemplo, aplicações baseadas nos conjuntos de dados produzidos pelo mesmo ED, ou um conjunto de serviços que faz uso de uma infraestrutura desenvolvida por outro ED. A saúde da rede de recursos pode ser avaliada de acordo com diversas métricas, por exemplo pelo nível de dependência entre os recursos, como eles podem impactar no pro-

cesso de criação de novas soluções, quais recursos são considerados prioritários naquele ED e como essa rede de recursos influencia nas decisões do ED.

*Saúde dos Papéis:* Os papéis são as funções que os atores estão exercendo naquele momento, podendo um ator exercer mais de um papel no mesmo ED ou em outros. Existem diversos papéis nos EDs, geralmente os mais relatados na literatura de ED são os consumidores e publicadores de dados. Para definir a saúde de cada papel é preciso definir os papéis existentes no ED em questão, avaliar as atividades, se há hierarquias nas tomadas de decisão, como são atribuídos os papéis e a influência que cada papel exerce sobre o outro.

Cada dimensão apresentada representa um aspecto válido que contribui para o funcionamento e gerenciamento do ecossistema. Após a formalização dos critérios de avaliação será feito o estudo de como cada critério será calculado e conseqüentemente realizar a operacionalização dos cálculos. Com isso serão obtidos indicadores que revelem o estado da saúde do ED e assim ajudar a identificar quais componentes ou atividades podem estar prejudicando o sucesso do ecossistema. E finalmente, para validar a estratégia proposta, serão realizados dois estudos de caso em EDs reais.

## 5. Método de Pesquisa

Este trabalho visa investigar e desenvolver uma estratégia de medição e avaliação da saúde de EDs. Nosso *design* de pesquisa pode ser definido como um paradigma de pesquisa pragmático que visa construir artefatos originais e inovadores para resolver problemas do mundo-real [Von Alan et al. 2004]. Isso envolve a criação e a avaliação de artefatos para resolver um problema específico. Baseados nos nossos objetivos e no paradigma pragmático, nós dividimos o design de pesquisa em três fases:

- A primeira fase visa mapear elementos e aspectos qualitativos e quantitativos citados na literatura que influenciam a saúde de EDs. Esses elementos e aspectos serão usados na construção de métricas que podem determinar a saúde de um ecossistema de dados.
- A segunda fase da pesquisa visa a formalização da estratégia de avaliação da saúde em EDs. Esta fase da pesquisa irá formalizar o problema a ser resolvido, definindo formalmente as métricas de avaliação, assim como a operacionalização do cálculo das métricas.
- A terceira fase da pesquisa consiste na validação da estratégia proposta. Serão realizados dois estudos de caso nos quais a estratégia será aplicada em dois EDs reais (ED da Universidade Federal de Pernambuco e o ED da cidade do Recife-PE) para avaliar a saúde de seus respectivos ecossistemas. E, posteriormente, esses participantes irão responder questionários de avaliação da estratégia com base em um conjunto de critérios de qualidade.

## 6. Trabalhos Futuros

Neste trabalho a necessidade de criar uma estratégia para avaliar a saúde de Ecossistemas de Dados foi apresentada. Enquanto trabalhos na área de ED salientam a necessidade de avaliar a saúde desses ecossistemas [Oliveira et al. 2017], nenhum trabalho com este propósito foi encontrado. As estratégias de avaliação da saúde de ecossistemas encontrados na literatura não são capazes de lidar com a heterogeneidade e dinamicidade dos componentes (atores, relacionamentos, recursos e papéis) presentes nos EDs.

Como próximos passos neste trabalho nós pretendemos: (i) finalizar o levantamento e especificação dos critérios e métricas para avaliação da saúde de EDs; (ii) formalizar a estratégia de avaliação de saúde em ecossistemas de dados; (iii) validar a estratégia proposta.

Como parte da pesquisa sobre Ecossistemas de Dados foi publicado o artigo "A Platform for Supporting Open Data Ecosystems" no ICEIS 2016. O plano para futuras publicações inclui: (i) "Investigations about data ecosystems: A systematic mapping study" submetido para o KAIS e (ii) "Investigations about Data on the Web Publication and Consumption: A Systematic Mapping Study" para o WWW journal.

## Referências

- Alves, C., Oliveira, J., and Jansen, S. (2017). Software ecosystems governance—a systematic literature review and research agenda. vol. 3:26–29.
- Costanza, R. (1992). Toward an operational definition of ecosystem health. *Ecosystem health: New goals for environmental management*. 239–256.
- den Hartigh, E., Tol, M., and Visscher, W. (2006). The health measurement of a business ecosystem. *Proceedings of the European Network on Chaos and Complexity Research and Management Practice Meeting*. 1–39.
- Iansiti, M. and Levien, R. (2004a). The keystone advantage: What the new dynamics of business ecosystems mean for strategy, innovation, and sustainability: Harvard business school press. *Boston, MA*.
- Iansiti, M. and Levien, R. (2004b). Strategy as ecology. *Harvard business review*, vol. 82(3):68–81.
- Lóscio, B. F., Burle, C., and Calegari, N. (2017). Data on the Web Best Practices. <https://www.w3.org/TR/dwbp/>. Acesso em: 01/05/2018.
- Manikas, K. and Hansen, K. M. (2013). Reviewing the health of software ecosystems—a conceptual framework proposal. *Proceedings of the 5th International Workshop on Software Ecosystems (IWSECO)*. 33–44.
- Oliveira, M. I. S., Lima, G. d. F. A. B., and Lóscio, B. F. (2017). Investigations about data ecosystems: A systematic mapping study. *Submetido para Knowledge and Information Systems*.
- Oliveira, M. I. S. and Lóscio, B. F. (2018). What is a data ecosystem? *Aceito em Digital Government Research*.
- Schaeffer, D. J., Herricks, E. E., and Kerster, H. W. (1988). Ecosystem health: I. measuring ecosystem health. *Environmental Management*, vol. 12(4):445–455.
- Von Alan, R. H., March, S. T., Park, J., and Ram, S. (2004). Design science in information systems research. *MIS quarterly*, vol. 28(1):75–105.
- Wahyudin, D., Mustofa, K., Schatten, A., Biff, S., and Min Tjoa, A. (2007). Monitoring the “health” status of open source web-engineering projects. *International Journal of Web Information Systems*, vol. 3(1/2):116–139.