

# MODELO DE MATURIDADE DE DADOS ABERTOS: UMA MATRIZ DE REFERÊNCIA PARA ORGANIZAÇÕES

Lidiane Visintin<sup>1</sup>, José Leomar Todesco<sup>1</sup>, Fernando Álvaro Ostuni Gauthier<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento  
Universidade Federal de Santa Catarina(UFSC)  
Florianópolis – SC – Brasil

lidiane.visintin@gmail.com, jose.todesco@ufsc.br, gauthier@egc.ufsc.br

**Abstract.** *Open data is a widespread topic due to the benefits that can be observed about them and also due to laws and regulations that some countries impose in search of transparency and other benefits. However, it is observed that there are organizations that make data available, but that are in the initial stages regarding practices and treatments for open data. In this sense, an open data maturity model is proposed, considering good practices applied by organizations in the opening and post-publication of data. To structure the model, Design Science Research is used. Based on this proposal, a reference matrix for maturity assessment was obtained, through which an ideal scenario of evolution for open data initiatives is envisaged, which can be restructured according to the reality of each organization.*

**Resumo.** *Dados abertos é um tema difundido devido aos benefícios que podem ser observados acerca destes e também devido a leis e regulamentações que alguns países impõem em busca de transparência e outros benefícios. No entanto, observa-se que há organizações que disponibilizam dados, mas que encontram-se em estágios iniciais com relação às práticas e tratamentos para dados abertos. Neste sentido, propõe-se um modelo de maturidade de dados abertos, considerando boas práticas aplicadas pelas organizações na abertura e na pós-publicação dos dados. Para a estruturação do modelo faz-se o uso da Design Science Research. Com base nesta proposta obteve-se uma matriz de referência para avaliação da maturidade, sendo que através desta vislumbra-se um cenário ideal de evolução para as iniciativas de dados abertos, sendo que esta poderá ser reestruturada conforme a realidade de cada organização.*

## 1. Introdução

A abertura de dados é amplamente difundida, compreendendo vários países e diferentes níveis administrativos [Huijboom and Van den Broek 2011]. Nesse cenário, as pessoas e organizações estão tendo acesso a dados de vários setores não só do contexto público. Observa-se este cenário através do número crescente de portais e de conjuntos de dados que estão acessíveis [Castellani Ribeiro et al. 2015, Tygel et al. 2016, McCrae and Abele 2021], são exemplos: DataPortals.org<sup>1</sup> que em maio de 2021 contabiliza cerca de 590 portais de dados abertos, considerando uma lista de portais acessíveis no mundo.

<sup>1</sup><http://datacatalogs.org/>

Dados abertos (DA) podem fornecer benefícios a várias áreas possibilitando impactos em setores como: educação, transporte, eletricidade, petróleo, saúde, e finanças [Manyika et al. 2013, Machado et al. 2019]. O impacto positivo pode ser alcançado devido as pessoas obterem e processarem mais facilmente os dados.

Os impactos positivos de DA podem ser observados na sociedade também, pois pode melhorar a eficiência das administrações públicas, aumenta o crescimento econômico no setor privado, criar novos produtos baseados em dados e contribuir para o espaço de formulação de políticas públicas [Kapoor et al. 2015]. No entanto, DA é um tema que necessita ser investigado, pois ainda há muitas organizações em um estágio inicial de abertura de dados, sendo que há barreiras e desafios relacionados a DA que precisam ser superados, para que as organizações obtenham sucesso em suas iniciativas de abertura de dados [Kaasenbrood et al. 2015, Wu and Mahdzir 2018, Husin et al. 2019].

Na literatura são encontrados diversos estudos que abordam diferentes desafios, obstáculos e problemas relacionados a DA, como: ausência de metadados, insuficiência de metadados, falta de interoperabilidade e dificuldade em pesquisar e navegar pelos dados [Tygel et al. 2016, Zuiderwijk et al. 2012]. Quando considerando os conjuntos de dados, também há indicativos de que existem vários problemas nos dados disponibilizados, incluindo: conformidade de metadados, duplicação, inconsistência e integridade de valores [Sadiq and Indulska 2017]. Toda estas inconsistências sobre os dados podem ocasionar mal-entendidos, dificuldade de acesso e dificuldade de processamento dos dados disponíveis. Além de que a complexidade de obter os dados influenciará a vontade, a capacidade e a intenção do usuário em fazer uso dos dados [Husin et al. 2019].

Neste sentido, surgiram iniciativas que utilizam de painéis e relatórios que se propõem a avaliar o engajamento das organizações na abertura de dados, destaca-se: Open Data Maturity<sup>2</sup>; O Modelo de Maturidade de DA de [Dodds and Newman 2015] e o Modelo de Maturidade de DA de [Solar et al. 2012], dentre outros. No entanto, nenhuma destas iniciativas tem como foco analisar e considerar as práticas e padrões adotados para dados, e assim surgiu a seguinte pergunta de pesquisa: “*Como avaliar a maturidade de dados abertos das organizações?*”.

Observando o cenário da pesquisa, optou-se pelo contexto da pesquisa tecnológica. Para isso fez-se o uso dos princípios da *design science research*, que por sua vez, é o método que fundamenta e operacionaliza a condução da pesquisa quando o objetivo a ser alcançado é um artefato [Dresch et al. 2015]. E neste âmbito, optou-se pelo método proposto por [Peffers et al. 2007], onde foram seguidas as seguintes etapas: (I) identificação do problema e motivação, (II) definir objetivos para a solução, (III) design e desenvolvimento, (IV) demonstração, (V) avaliação e (VI) comunicação dos resultados.

## 2. Modelo de Maturidade de Dados Abertos

O domínio considerado para a estruturação do modelo de maturidade é o da abertura e de pós publicação de DA, ou seja, foi considerado o ciclo de vida de DA. As áreas foco, denominadas de: (I) Gerenciamento de políticas e processos, (II) Gerenciamento de fatores organizacionais e (III) Gerenciamento de padrões foram definidas com base na governança de dados, com o objetivo de lidar com problemas de dados e garantir que

<sup>2</sup>Disponível em: <https://data.europa.eu/en/dashboard/2020>

os dados atendam as necessidades dos negócios, sendo considerado o *framework* definido por [Panian 2010], para evidenciar estas áreas. Na Tabela 1 apresenta-se a estrutura elaborada.

**Tabela 1. Matriz - Modelo de Maturidade de Dados Abertos**

Níveis de Maturidade	0	1	2	3	4	5	6	7	8	9	10	11	12
<b>Gerenciamento de políticas e processos</b>													
Definição dos dados				A			B			C			D
Acesso e entrega de dados				A		B		C		D			
Monitoramento e medição			A	B	C	D							
Gerenciamento de mudança de dados				A			B		C		D		
<b>Gerenciamento de fatores organizacionais</b>													
Funções e responsabilidade			A	B	C	D							
Educação		A	B	C	D								
Treinamento			A	B	C		D						
Planejamento e priorização da abertura dos dados					A		B		C		D		
Planejamento da publicação dos dados						A		B			C	D	
Fomento ao uso		A		B		C		D					
<b>Gerenciamento de padrões</b>													
Licença		A		B	C		D						
Proveniência e qualidade		A	B	C			D						
Versionamento		A	B	C		D							
Identificação		A		B	C	D							
Formato					A	B	C				D		
Vocabulários		A	B	C					D				
Metadados		A		B	C				D				
Preservação								A	B	C	D		
Enriquecimento		A	B	C	D								
Feedback		A		B	C	D							
Acesso								A		B	C		D

Fonte: Da Autora

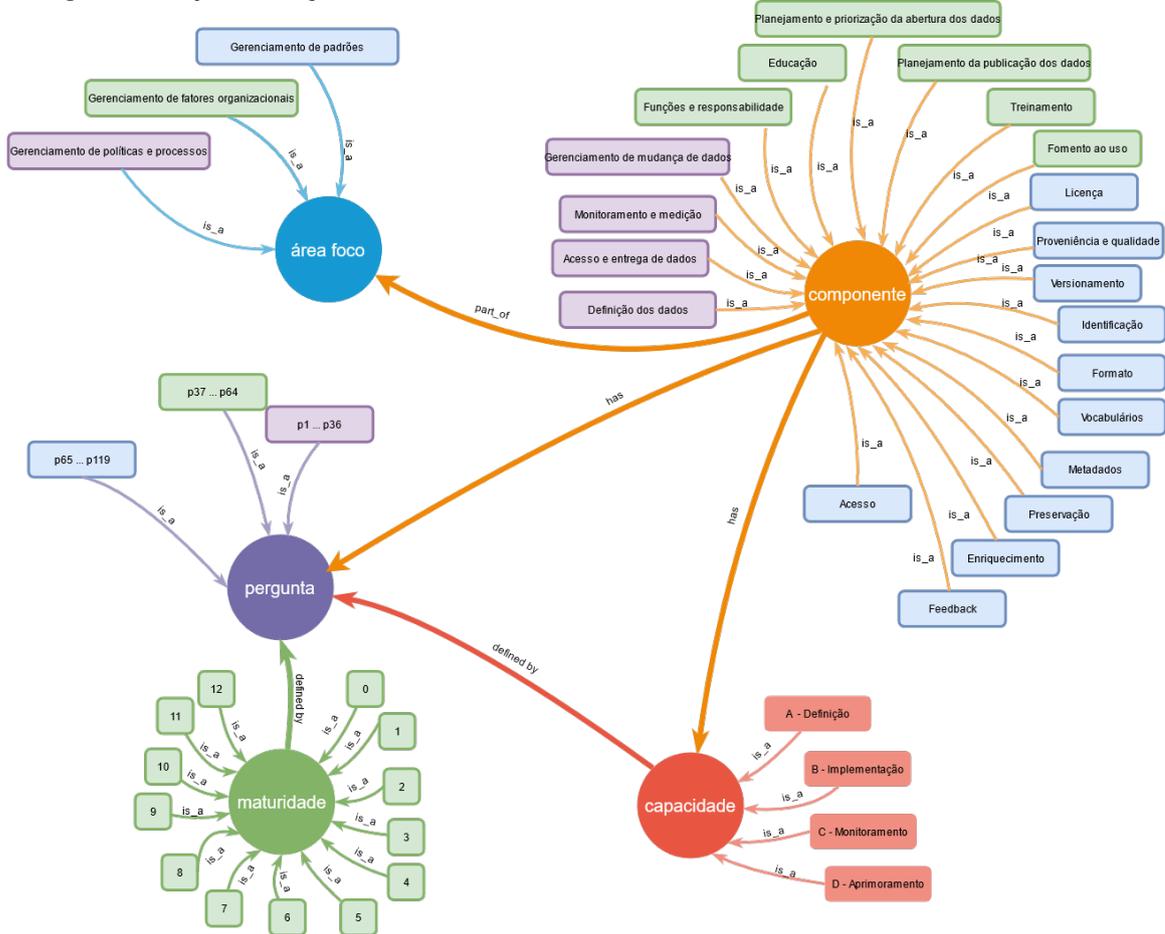
Para cada área foco foram apresentados os elementos que a compõe, estes denominados de “componente”. Sendo que para as áreas foco Gerenciamento de políticas e processos, Gerenciamento de fatores organizacionais foram considerados os elementos apresentados no trabalho de [Panian 2010], já para a área foco Gerenciamento de padrões foram considerados os desafios apresentados em [Lóscio et al. 2017].

As capacidades foram elencadas com base em três ações que indicam a execução da governança de dados, sendo estas: A - definir, B - implementar e C - monitorar [Alhassan et al. 2016]. Uma 4ª atividade denominada por D - “Aprimorar” será utilizada, com o intuito de garantir a continuidade no fornecimento de DA, com no mínimo a mesma qualidade e com a possibilidade de evoluir as políticas, processos e padrões.

Após a definição das capacidades, estas foram posicionadas na matriz, de acordo com as dependências observadas para cada componente e seguindo os princípios da

estruturação de modelos de maturidade de áreas foco [Van Steenbergen et al. 2010]. A maturidade sempre tem início na Coluna 1, pois a maturidade 0 corresponde a inexistência de práticas relacionadas ao componente.

**Figura 1. Representação do conhecimento abordado no modelo de maturidade**



Fonte: Da autora, 2021

A Figura 1 apresenta a estrutura desenvolvida baseada na representação utilizada em ontologias, com o intuito de abordar e sumarizar os conhecimentos observados no decorrer desta pesquisa e suas inter-relações. Sendo possível através desta figura observar e compreender as relações existentes entre as áreas foco, os componentes e as perguntas para com as capacidades e os níveis de maturidade.

Com o intuito de observar o comportamento do modelo foram entrevistas para coletar cenários de uso reais, e também fez-se a avaliação da estrutura do modelo, considerando a opinião de especialistas da área. Com base nas entrevistas realizadas, foi possível observar que o modelo de maturidade possibilita avaliar a maturidade de organizações que realizam a abertura de dados e esta pesquisa pode servir de referência para possíveis desdobramentos em novos modelos, de modo a atender cenários em específico.

### 3. Considerações Finais

Todos os esforços da pesquisa realizada se deram com a intenção de se ter atividades no modelo que forneçam um roteiro para melhorias e não uma lista de tarefas para serem

completadas antes dos dados serem liberados. De fato, é desejável que as organizações experimentem baixos níveis de maturidade para que desenvolvam processos adequados para si próprias. Pois, a tentativa de saltar para níveis mais elevados de maturidade pode resultar em processos impostos que não levam em conta a cultura, o ambiente ou as necessidades específicas das organizações. Assim, a proposição do artefato também se dá no sentido de obter uma estrutura que não sirva como fator de desmotivação para as organizações, mas para que se tenha uma estrutura adequada as diferentes realidades.

Para a análise das diferentes realidades foi desenvolvida uma aplicação Web<sup>3</sup>, para a coleta dos cenários de uso, nesta página há informações sobre a pesquisa que está sendo desenvolvida além de permitir responder ao questionário de avaliação da maturidade e com isso visualizar a matriz de maturidade conforme as respostas fornecidas.

Neste sentido, salienta-se que o modelo de maturidade organizado em áreas foco, oriundas da governança de dados e componentes viabilizam a avaliação da maturidade de modo a aprimorar cada componente de interesse da organização, tornando a estrutura do artefato adaptável aos diferentes cenários observados para DA. Sendo que, o trabalho encontra-se em fase de finalização e que possibilitará a elaboração de estudos de casos e em cenários similares ou até mesmo em níveis de governo, isso como trabalhos futuros, para compreender a estrutura necessária para a avaliação da maturidade.

Por fim, destacam-se as contribuições da Engenharia do Conhecimento na realização desta pesquisa, pois ao final da concepção do modelo de maturidade foi possível se ter uma representação dos conhecimentos abordados, considerando como base alguns dos princípios de ontologias e desta maneira observa-se que podem ser derivados novos estudos no âmbito semântico e computacional.

## Referências

- Alhassan, I., Sammon, D., and Daly, M. (2016). Data governance activities: an analysis of the literature. *Journal of Decision Systems*, 25(sup1):64–75.
- Castellani Ribeiro, D., Vo, H. T., Freire, J., and Silva, C. T. (2015). An urban data profiler. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pages 1389–1394, New York, NY, USA. ACM.
- Dodds, L. and Newman, A. (2015). *Guia do Modelo de Maturidade de Dados Abertos*. Open Data Institute.
- Dresch, A., Lacerda, D. P., and Júnior, J. A. V. A. (2015). *Design science research: método de pesquisa para avanço da ciência e tecnologia*. Bookman Editora.
- Huijboom, N. and Van den Broek, T. (2011). Open data: an international comparison of strategies. *European journal of ePractice*, 12(1):4–16.
- Husin, N. N. F. A., Zakaria, N. H., and Dahlan, H. M. (2019). Factors influencing open data adoption in malaysia based on users perspective. In *2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS)*, pages 1–5.
- Kaasenbrood, M., Zuiderwijk, A., Janssen, M., de Jong, M., and Bharosa, N. (2015). Exploring the factors influencing the adoption of open government data by private orga-

<sup>3</sup>disponível em: <https://lidianevisintin.studiokem.ufsc.br/>, sendo que há controle de acesso ao questionário, devido a ser um estudo em andamento e que possivelmente necessitará de ajustes;

- nisations. *International Journal of Public Administration in the Digital Age (IJPADA)*, 2(2):75–92.
- Kapoor, K., Weerakkody, V., and Sivarajah, U. (2015). Open data platforms and their usability: Proposing a framework for evaluating citizen intentions. In *Conference on e-Business, e-Services and e-Society*, pages 261–271. Springer.
- Lóscio, B. F., Burle, C., and Calegari, N. (2017). Data on the web best practices. <https://www.w3.org/TR/dwbp/>.
- Machado, L. M. O., Souza, R. R., and da Graça Simões, M. (2019). Semantic web or web of data? a diachronic study (1999 to 2017) of the publications of tim berners-lee and the world wide web consortium. *Journal of the Association for Information Science and Technology*, 70(7):701–714.
- Manyika, J., Chui, M., Farrell, D., Kuiken, S. V., Groves, P., and Doshi, E. A. (2013). Open data: Unlocking innovation and performance with liquid information. <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>.
- McCrae, J. P. and Abele, A. (2021). The linked open data cloud. <https://lod-cloud.net/>. 21/04/2021.
- Panian, Z. (2010). Some practical experiences in data governance. *World Academy of Science, Engineering and Technology*, 62(1):939–946.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77.
- Sadiq, S. and Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management*, 37(3):150–154.
- Solar, M., Concha, G., and Meijueiro, L. (2012). A model to assess open government data in public agencies. In *International Conference on Electronic Government*, pages 210–221. Springer.
- Tygel, A., Auer, S., Debattista, J., Orlandi, F., and Campos, M. L. M. (2016). Towards cleaning-up open data portals: A metadata reconciliation approach. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, pages 71–78.
- Van Steenbergen, M., Bos, R., Brinkkemper, S., Van De Weerd, I., and Bekkers, W. (2010). The design of focus area maturity models. In *International Conference on Design Science Research in Information Systems*, pages 317–332. Springer.
- Wu, T. and Mahdzir, A. M. (2018). An investigation of application developer’s awareness and attitude towards open government data in malaysia. *International Journal of Engineering & Technology*, 7(4):5127–5131.
- Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., and Alibaks, R. S. (2012). Socio-technical impediments of open data. *Electronic Journal of e-Government*, 10(2).