

Modelo de infraestrutura para publicação de dados abertos educacionais conectados de qualidade

Bruno Elias Penteado, José C. Maldonado (coorientador), Seiji Isotani (orientador)

¹Instituto de Ciências Matemáticas e da Computação – Universidade de São Paulo
(ICMC/USP) – São Carlos – SP – Brasil
brunopenteado@alumni.usp.br, {jcmaldon, sisotani}@icmc.usp.br

***Resumo.** Os dados abertos disponibilizados por governos em todo o mundo têm sido usados para o desenvolvimento de pesquisas, serviços e políticas públicas, gerando insights e embasando ações. Um desafio atual é a dificuldade em conectar diferentes fontes desses dados para realizar cruzamentos complexos, devido a problemas como interoperabilidade e alinhamento semântico. Esta tese propôs um arcabouço de artefatos que subsidiam a publicação de dados abertos governamentais educacionais, dotados de mecanismos de controle de qualidade e de acordo com as melhores práticas de compartilhamento de dados na Web. Foram desenvolvidos três artefatos tecnológicos, sob a epistemologia da Design Science Research. Juntos, triangulam para tratar o problema de modo mais completo do ponto de vista tecnológico. Foram conduzidos i) um estudo de caso sobre o problema da rotatividade docente, unindo diferentes fontes de dados e visualizando as movimentações e ii) um quasi-experimento em que os sujeitos publicaram dados educacionais de forma consistente e seguindo as melhores práticas em qualidade de dados conectados e compartilhamento de dados. Esta tese avança sobre a teoria do design em publicação de dados educacionais, desenvolve uma infraestrutura de publicação, modelo de referência semântico para dados educacionais e demonstra sua aplicabilidade em problemas reais.*

1. Introdução

A produção de dados abertos governamentais (DAG) tem aumentado no mundo todo, viabilizando a transparência de governos e considerada como um dos pilares da democracia moderna. A disponibilidade desses dados traz promessas como: o desenvolvimento econômico, o controle social, o embasamento de políticas públicas, o combate à corrupção, dentre outras. Diferentes órgãos governamentais divulgam informações de seu escopo, tornando possível que consumidores de dados combinem diferentes fontes para responder questões que antes não eram possíveis. No Brasil, diversas pesquisas têm sido elaboradas baseadas em dados do Censo Escolar, IDEB, ENEM, dentre outros, para o acompanhamento de políticas públicas em diferentes níveis.

Embora tenha havido progressos variados em torno desses eixos, vários fatores impedem que esses benefícios sejam realizados em maior escala. Da perspectiva tecnológica, a disponibilização dos dados que possam ser combinados, sua formatação padronizada e sua interoperabilidade são talvez os fatores que mais impactam sua publicação. Ademais, as tecnologias semânticas têm sido desenvolvidas para tratar este

problema de interoperabilidade ao representar informações de modo que possam ser interpretadas inequivocamente por humanos e máquinas, por meio dos chamados *dados conectados* [Isotani & Bittencourt, 2015]. No entanto, as metodologias existentes apresentam limitações como a percepção de serem muito genéricas e de não tratarem corretamente questões de qualidade dos dados¹, fazendo com que publicadores adotem métodos *ad-hoc* de publicação, apresentando resultados inconsistentes.

Assim, o objetivo geral desta tese foi *desenvolver uma infraestrutura de publicação de dados abertos governamentais conectados (DAGC) de qualidade para o ecossistema de dados da Web*, com aplicação na área educacional. Por infraestrutura, entendemos o conjunto de artefatos necessário para atender o problema adequadamente.

Dada esta visão geral, os seguintes objetivos específicos foram definidos:

- **Objetivos específicos:** (OE1) descrever como as iniciativas de dados abertos evoluíram em nível nacional, ou seja, o estado da prática; (OE2) identificar o estado da arte de como os processos atuais operacionalizam a publicação dos DAGC; (OE3) propor uma arquitetura baseada nos paradigmas e requisitos da Web como meio de compartilhamento; (OE4) especificar detalhadamente as fases, atividades, ferramentas e papéis necessários para o processo de publicação de DAGC; (OE5); validar os artefatos propostos a partir de dados educacionais reais.

A relevância deste trabalho para a comunidade de Informática na Educação dá-se por sua aplicabilidade no setor educacional - demonstrada na validação dos artefatos, ao fornecer mecanismos para conectar dados de diferentes fontes e viabilizar novos estudos a partir dos dados publicados. A originalidade do tema vem da importância com que a Web se transforma em plataforma de publicação, compartilhamento e conexão de dados para a geração de conhecimento, muitas vezes distribuído em diferentes locais e formatos, bem como quais novos tipos de análises educacionais podem ser potencializados com a aplicação dessas técnicas. As áreas de Educação e Informática na Educação têm muito a ganhar com essa abordagem, pois ao superar os desafios técnicos, possibilita a descoberta de novos conhecimentos para problemas complexos existentes – tanto educacionais como tecnológicos, tendo a Web e agentes computacionais como componentes desta solução.

Esta tese traz contribuições em diferentes dimensões:

Teórica: avança sobre a teoria do *design* [Baskerville & Pries-Heje, 2010] em publicação de dados educacionais, ao apresentar requisitos levantados da comunidade de prática e ao abordar teorias *kernel* que os implementam; desenvolve uma infraestrutura de publicação com processo, arquitetura e metadados, além de um modelo de referência semântico para dados educacionais.

Metodológica: maior rigor avaliativo do que o presente no estado da arte.

¹ Qualidade de dados é um termo que pode contemplar diferentes dimensões. Nesta tese, a qualidade diz respeito às características únicas dos dados conectados, com por exemplo: representação dos dados, consistência em relação à informação apresentada em outros datasets publicados, qualidade da representação semântica, metadados semânticos e a acurácia das conexões feitas.

Prática: analisa historicamente e criticamente os dados governamentais educacionais; pormenoriza as atividades, ferramentas e checagens a serem feitas em futuras publicações; demonstra sua aplicabilidade em problemas reais de educação.

2. Trabalhos relacionados, estado da arte e da prática

O ecossistema de DAG é um tema recente de pesquisa, com um corpo de conhecimento ainda em formação [Oliveira et al., 2019]. Diversos autores exploraram o foco educacional: Santos (2014) compara o uso dos DAG no Brasil e no Reino Unido; Bandeira et al. (2014) trazem os potenciais benefícios e limitações do uso de dados conectados na educação, como o custo de publicação, o reuso de vocabulários e a qualidade dos dados; Pereira et al. (2018) mapeiam ontologias, vocabulários e ferramentas neste contexto, apontando como um dos desafios a diversidade de vocabulários e ferramentas adotadas.

De forma mais ampla, outros autores projetaram metodologias para a publicação de DAGC [Auer et al., 2011; Villazón-Terrazas et al., 2011] e até mesmo modelos de referência para sistemas de gestão de dados conectados [Oliveira et al., 2018]. No entanto, tais recomendações e metodologias são muitas vezes consideradas genéricas [Laessig et al., 2019], sem detalhar quais passos e ferramentas devem ser utilizados. Como resultado, a maioria dos estudos que produzem dados conectados não especifica o uso de metodologias sistemáticas, optando por abordagem *ad-hoc* [Santos et al., 2018].

Por se tratar de um assunto multidisciplinar, optou-se por revisar o estado da prática, além do estado da arte presente na literatura científica, para melhor compreender quais fenômenos influenciaram sua evolução, incluindo a perspectiva tecnológica.

Para o estado da prática (OE1), realizamos um estudo historiográfico mapeando os principais marcos relacionados à publicação de DAG no Brasil, desde a Constituição de 1988 até a Política Nacional de Dados Abertos de 2019, relacionando-os aos principais marcos em diferentes países. Pode-se notar que o Brasil esteve na vanguarda, junto a países desenvolvidos. Porém, ainda não existem dados conectados produzidos em larga escala. Em outro estudo [Penteado et al. 2019a], focamos nos DAG educacionais, analisando seu grau de atendimento às melhores práticas de compartilhamento de dados, como preconizado pela W3C [W3C, 2017]. Notou-se que a maior parte dessas práticas não era plenamente atendida, inferindo-se que a publicação foi otimizada para acesso humano e não para o processamento na Web semântica.

Para compreender o estado da arte (OE2), realizamos um mapeamento sistemático com 19 estudos mapeados que propuseram metodologias de publicação de DAGC, independente de área. Na síntese, foram mapeados: i) os passos comuns entre as metodologias, ii) as ferramentas de suporte utilizadas, iii) as formas de avaliação e iv) a existência de mecanismos de controle de qualidade. Como resultados, notou-se um conjunto comum de passos e ferramentas, além de outros passos contemplando diferentes dimensões do processo. Descobriu-se também que as metodologias foram fracamente avaliadas, majoritariamente com cenários ilustrativos. Por fim, poucos mecanismos de controle de qualidade foram abordados. Estas podem ser as causas de que as metodologias

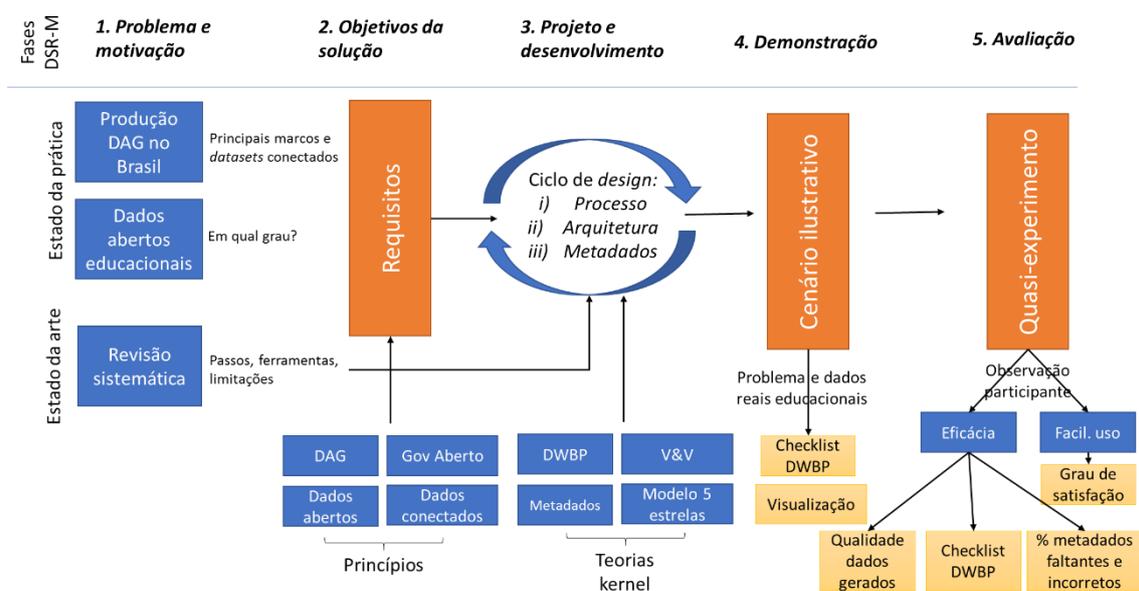
são vistas como genéricas para publicadores de dados [Penteado et al., 2021]. Esses resultados balizaram a construção dos artefatos da tese, estendendo o estado da arte com contribuições nestes aspectos. A partir destes mapeamentos do estado da prática no Brasil, do estado da arte na literatura científica de DAGC e de trabalhos na Informática na Educação esta tese foi posicionada e desenvolvida.

3. Procedimento metodológico

A metodologia de pesquisa adotada nesta tese foi a DSR-M (*Design Science Research Methodology*; [PEFFERS et al., 2007]), baseada na epistemologia da *Design Science Research*, que busca desenvolver artefatos para compreender e intervir no mundo natural.

A DSR-M contém 6 passos sequenciais e iterativos, conforme especificado na Figura 1. O último passo (Comunicação) tem por resultado esta tese e seus artigos.

Figura 1. Fases da pesquisa da tese, seguindo a metodologia DSR-M.



Para definir os requisitos da infraestrutura proposta na tese, reunimos os princípios dos dados abertos, governo aberto, dados abertos governamentais, dos dados conectados, transformando-os em uma lista com requisitos disjuntos entre si. De modo a se chegar a uma versão refinada dos artefatos, foi selecionado um cenário ilustrativo, em que é feita a aplicação dos artefatos a uma situação do mundo real, objetivando ilustrar sua aptidão ou utilidade. Este método de avaliação é indicado para o refinamento de artefatos na DSR, ao buscar replicar um contexto ideal para a aplicação do artefato. Por meio deste cenário, os artefatos foram sendo refinados iterativamente, em sucessivas tentativas de design. Ao fim, chegou-se ao conjunto de artefatos discutidos na seção 4.1. Seu produto foi avaliado por meio de checklist em função das melhores práticas da W3C no compartilhamento de dados pela Web [W3C, 2017].

Em seguida, os artefatos projetados e refinados foram submetidos ao uso por 5 sujeitos, em um *quasi*-experimento – sem condição de controle e com pós-teste – em que os participantes possuíam um roteiro para publicação dos dados, com o apoio da

infraestrutura proposta na tese. Como o ecossistema de dados abertos é um campo de pesquisa recente, não existem ainda teorias específicas a serem testadas, mas sim desenvolvidas, daí a opção por esta escolha. A participação foi remota, com acompanhamento pelo autor da tese via Skype. Para cada artefato, foi avaliada sua eficácia (atendimento aos requisitos) e sua facilidade de uso. A eficácia foi operacionalizada como a seguir: i) as métricas de qualidade dos dados gerados pela execução do processo, por meio do arcabouço de métricas de dados conectados proposto por Debattisti et al. (2018). Três *datasets* em formato conectado foram selecionados como padrão-ouro para a comparação dos dados gerados pelos participantes: dados do Orçamento Federal (LOA), Previdência Social (AEPS) e Banco Mundial (WB); ii) o número de requisitos atendidos da DWBP [W3C, 2017] em relação à arquitetura e ambientes instanciados, iii) a proporção de metadados faltantes ou incorretos em relação ao total requerido em um *template*. Um cenário simples foi apresentado aos participantes, especificado por meio de um roteiro e propondo a conexão de dados educacionais de 3 fontes diferentes. A facilidade de uso foi medida por meio de questionários aplicados aos participantes, indagando sobre a facilidade do uso de cada artefato via itens Likert, bem como pelo acompanhamento por meio de observação indireta e participante.

4. Resultados

Esta seção traz os artefatos resultantes, conforme previsto pela DSR e suas avaliações em relação ao tratamento dos problemas mapeados (OE3 e OE4).

4.1. Artefatos

Foram selecionados corpos de conhecimento que atendessem aos requisitos levantados, formando um embasamento teórico para a criação dos artefatos. Foram selecionados i) o *ecossistema de dados abertos*, concretizado pelo conjunto de melhores práticas para o compartilhamento de dados na Web, da W3C (DWBP), ii) *controle de qualidade*, por meio de mecanismos de verificação e validação (V&V), iii) o *modelo de 5 estrelas*, de requisitos técnicos para os dados conectados e iv) o corpo de conhecimento sobre *metadados* e suas técnicas para indexação e recuperação de informações, tanto por humanos quanto por máquinas. A partir desses materiais, foram criados i) um *modelo de ciclo de vida*, com modelo de processo com os passos comuns identificados no estado da arte e estendidos para atender a todos os requisitos levantados; ii) uma arquitetura conceitual, especificando as camadas de integração, os pontos de controle de qualidade e o encaixe de cada componente necessário para dar suporte à infraestrutura concebida; iii) um conjunto mínimo de *metadados*, para que tanto humanos quanto agentes de software possam localizar os pontos de acesso aos dados, seus formatos, protocolos aceitos, etc. Os artefatos são ilustrados na Figura 2.

4.2. Avaliação empírica

A validação dos artefatos se deu em dois momentos (OE5). Como descrito na metodologia, foi conduzida inicialmente a avaliação por cenário ilustrativo. Foi selecionado um problema educacional relevante: a rotatividade docente nas escolas

públicas brasileiras - fenômeno que afeta significativamente a aprendizagem dos alunos. Foram usados dados do Censo da Educação Básica dos anos 2016 e 2017, conectados a dados do município de São Paulo. Os dados foram publicados conforme o modelo de processo e metadados propostos e distribuídos conforme a arquitetura. Foi criada uma visualização de dados (Figura 3), já que este é um problema complexo para acompanhamento. Por meio dos resultados, foi possível identificar movimentações fora do comum, por meio do cruzamento de diferentes fontes de dados, estendendo a literatura educacional, uma vez que cruzamentos desses dados não haviam sido feitos para este problema – limitados apenas a tabelas com estatísticas descritivas [Penteado & Isotani, 2021]. A arquitetura atendeu plenamente 25 e parcialmente 8 dos 35 requisitos propostos.

Com os artefatos refinados por meio do cenário ilustrativo, foi conduzido o *quasi*-experimento, em que tanto o método em si quanto os produtos foram avaliados.

Cada artefato foi validado conforme as métricas especificadas na metodologia. O processo, analisado conforme os dados gerados pelos participantes, teve o resultado conforme a Figura 4. Foi feita a média aritmética dos escores dos participantes em cada métrica para a comparação com os *datasets* padrão-ouro. Das 19 métricas analisadas, a média produzida pelos participantes foi maior ou igual em 15 delas. Algumas das métricas (ex: P1, P2, U1 e U5) sequer foram atendidas pelos *datasets* padrão-ouro, o que demonstra que nosso processo gera maior cobertura em relação a essas métricas. Em relação aos 35 requisitos da DWBP, em média, 29 foram plenamente atendidos, 4 parcialmente atendidos e 2 não foram aplicáveis ao cenário, com variação muito pequena entre os participantes. Embora diferentes ferramentas e ambientes tivessem de ser utilizados para a instanciação da arquitetura, os participantes o fizeram sem grandes problemas. Quanto aos metadados, foram detectados ~5% faltantes e 1,5% incorretos, dentre cem metadados a serem preenchidos por cada participante. A existência e a acurácia desses metadados estavam medidos dentre os requisitos da DWBP. No entanto, esta foi a principal reclamação por parte dos participantes, pois não existem ferramentas apropriadas para isto e todos desconheciam a importância dos metadados.

5. Discussão

O processo de publicação de DAGC mostrou-se uma atividade complexa, que aborda diferentes tarefas sociotécnicas para sua apropriada execução. Nesta tese, foi desenvolvida uma infraestrutura composta de 3 artefatos para dar suporte a esta tarefa, pormenorizando cada componente e suas ferramentas de suporte. Os dados do *quasi*-experimento demonstraram a viabilidade da solução; isso aliado a uma pequena variância entre os participantes, demonstrando a consistência e reprodutibilidade do processo proposto. Embora exista consenso sobre as melhores práticas para compartilhar dados na Web, não havia uma representação de como se conseguir isso. Por meio da arquitetura conceitual proposta, conseguimos atender a estes requisitos unindo diferentes perspectivas integradas entre si. Os metadados são geralmente relegados, deixando para a interpretação humana o significado dos dados. Para obter uma exploração automática dos dados, foi proposto um conjunto de metadados.

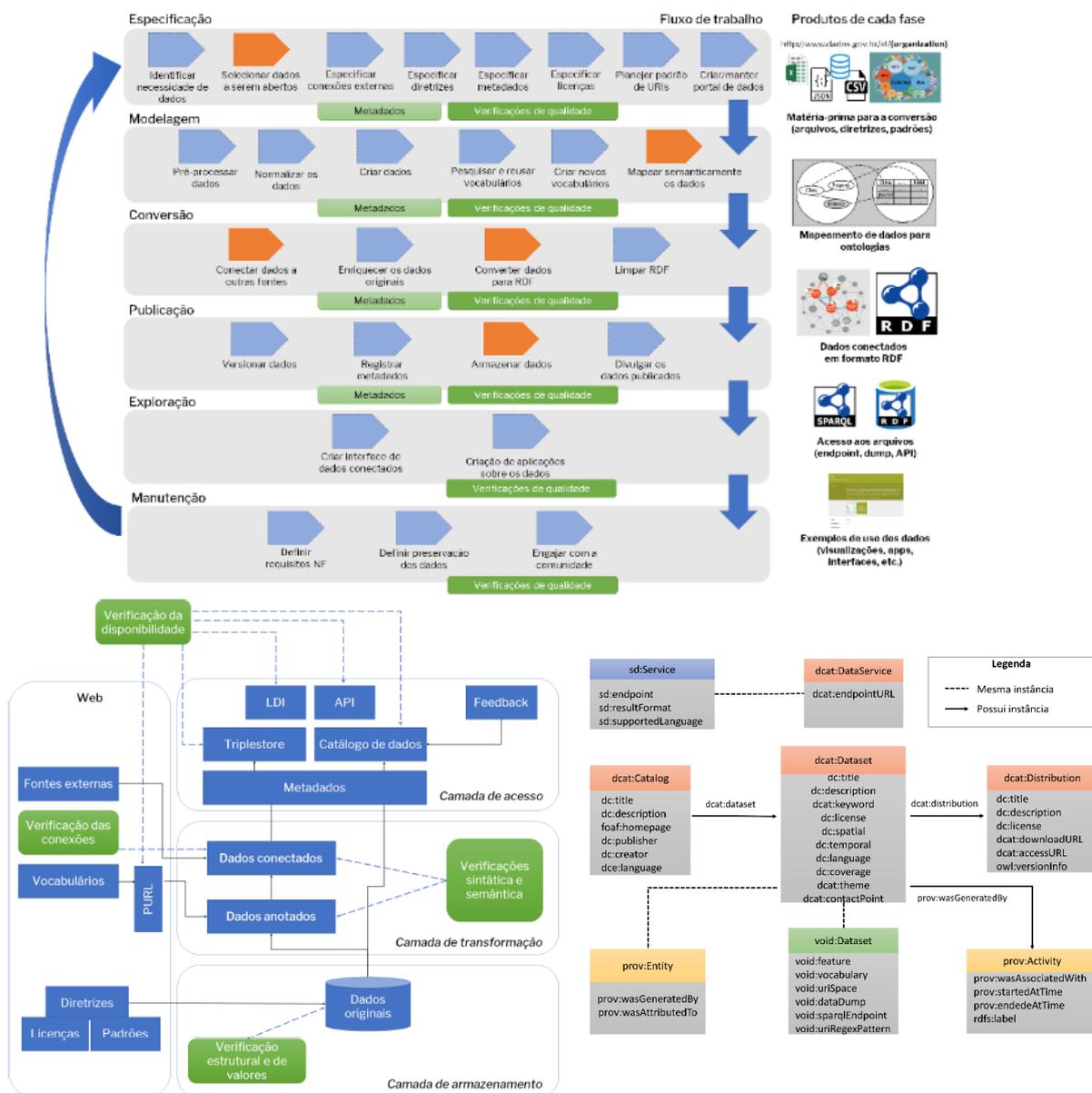


Figura 2. Artefatos gerados pela tese. Respectivamente: modelo de processo (fases, atividades e produtos), arquitetura (camadas, componentes e pontos de controle) e metadados (vocabulários e propriedades).

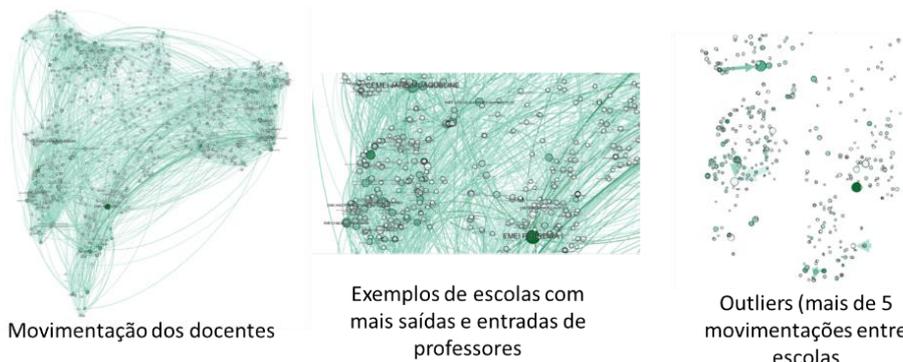


Figura 3. Resultado com a visualização das movimentações dos docentes.

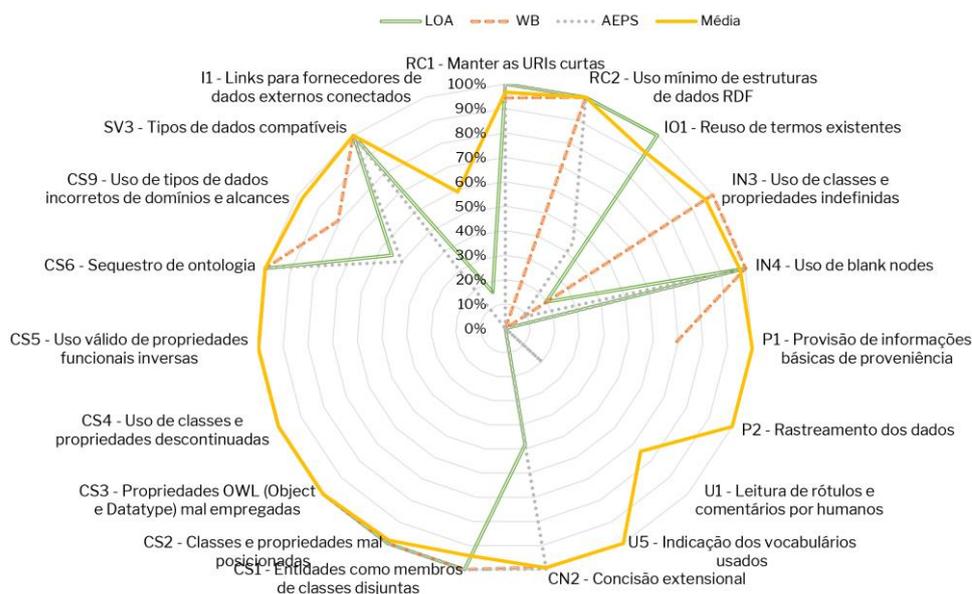


Figura 4. Comparação da média das métricas geradas pelos participantes (em amarelo) em relação aos datasets padrão-ouro.

A relevância da tese pôde ser observada pelo cenário ilustrativo utilizado no refinamento dos artefatos. Um problema relevante e complexo do domínio educacional foi visualizado de uma forma diferente, possibilitando novos *insights* sobre o problema. Outras possibilidades, como as inferências semânticas, podem ser exploradas de modo a se tirar mais proveito dos dados conectados e gerar novos conhecimentos. Novas ferramentas também são necessárias para suportar melhor a infraestrutura proposta. Este estudo mostra ser o mais detalhado dentre a literatura levantada, abordando desde questões conceituais, teoria do design, os detalhamentos de cada atividade e como estas se encaixam no ecossistema de dados abertos. Novos artigos ainda devem ser publicados com os resultados.

A tese estendeu o estado da arte ao propor os artefatos necessários - e não apenas um processo - para dar suporte a esta publicação, além de validá-los empiricamente – o que a revisão da literatura demonstrou não ter sido apropriadamente feito. Modelos de publicação de DAGC prévios desconsideram atividades que contemplem requisitos de controle de qualidade e de garantia de atendimento às melhores práticas de compartilhamento de dados na Web. Ademais, foi feito o mapeamento do que já foi desenvolvido em educação e o apontamento de oportunidades em sua aplicação.

Esta tese traz algumas limitações. A escolha do *quasi*-experimento foi feita considerando a complexidade da experimentação de um processo longo como o aqui proposto. Um estudo de caso longitudinal com órgão governamental é uma estratégia de trabalho futuro para aferir melhor sua adequação a um cenário com as contingências desta atividade. A suposição da qualidade de dados conectados traz apenas um recorte desta área de pesquisa, mais relacionada à estrutura da representação do dado semântico. Trabalhos futuros podem explorar melhor outras dimensões de qualidade. O modelo de processo proposto assemelha-se a um paradigma em cascata, com sequencias ordenadas

e bem-definidas. Princípios de metodologias ágeis podem ser aplicados a esta modelo, de modo a torná-lo mais responsivo a mudanças de requisitos na publicação de dados.

6. Conclusão

Com base nos resultados, pudemos inferir que objetivos da tese foram atendidos. O desenvolvimento da infraestrutura abordou seu contexto histórico e adicionou práticas inovadoras ao processo de publicação de DAGC. O cenário ilustrativo adotado, de alta relevância para a comunidade de educação, permitiu demonstrar o potencial para estender as análises feitas até então na literatura educacional, facilitando a visualização de dados complexos (espaciais e temporais) e a detecção de padrões que seriam de grande dificuldade se não por meio dos dados gerados pelos artefatos desenvolvidos. Além disso, o conjunto de artefatos também se mostrou importante para a publicação consistente de DAGC, demonstrado pelo *quasi*-experimento com diferentes participantes, que geraram dados educacionais de alta qualidade, comparáveis aos datasets considerados como padrão-ouro pela comunidade de dados conectados. Como trabalhos futuros, espera-se desenvolver ferramentas mais adequadas para a instanciação desta solução – o que ainda é insuficiente conforme levantado ao longo dos estudos. Além disso, é necessário um maior rigor na avaliação, particularmente com o estudo de caso aplicado em alguma organização e com usuários experientes no tópico, para investigar sua aderência prática.

7. Contribuições do trabalho para a área de Informática na Educação

Como demonstrado pelas publicações nas áreas de educação e informática educacional, podemos destacar as seguintes contribuições teóricas, metodológicas e técnicas:

- Mapeamento histórico da produção de dados abertos no Brasil, em particular no âmbito educacional, analisando seu grau de abertura [Penteado et al. 2017; Penteado 2019 – prêmio de melhor artigo de sua trilha no SBIE e submetido como artigo completo na RBIE] e a cobertura desses dados [Penteado & Isotani, 2017];
- Revisão dos processos de publicação de DAGC, mapeando suas fases, atividades, ferramentas, controles de qualidade e métodos de validação [Penteado et al, 2021].
- Modelo de referência para a modelagem semântica de dados educacionais em nível macro (sistemas de ensino) [Penteado et al, 2019b];
- Demonstração da viabilidade dos dados conectados publicados como forma de obter *insights*, como o impacto de fatores socioeconômicos [Penteado, 2016] e a visualização da rotatividade docente [Penteado et al 2020; Penteado & Isotani, 2021].
- Desenvolvimento e validação de conjunto de artefatos validados para a produção de DAGC: processo [Penteado et al., 2020], arquitetura e metadados.
- Validação do processo por meio dos principais cenários educacionais que utilizam dados conectados e um racional para a seleção de ferramentas [Penteado et al, 2019c].

Referências

Auer, S.; Lehmann, J.; Ngomo, A.C.N. (2011). Introduction to linked data and its lifecycle on the web. Reasoning web: semantic technologies for the web of data, p. 1-90. Springer, 2011.

- Bandeira, J., Alcantara, W., Barbosa, A., Ávila, T., Bittencourt, I., Isotani, S. (2014). Dados Abertos Conectados na Educação Brasileira. Jornada de Atualização em Tecnologia da Informação. Simpósio Brasileiro de Tecnologia da Informação.
- Baskerville, R., & Pries-Heje, J. (2010). Explanatory design theory. *Business & Information Systems Engineering*, 2(5), 271-282.
- Debattista, J.; Lange, C.; Auer, S. (2018). Evaluating the quality of the LOD Cloud: An empirical investigation. *Semantic Web Journal*, v. 9, n. 9, p. 859–901.
- Isotani, S.; Bittencourt, I. I. Dados abertos conectados. São Paulo: Novatec, 2015. 175 p.
- Laessig, M; Jacob, B.; AbouZahr, C. (2019). Opening data for global health. *The Handbook of Global Health Data Methods for Policy and Practice*. Springer Berlin, 2019, ch. 23, p.451-468.
- Oliveira, L.; Oliveira, M.; Santos, W.; Lóscio, B. (2018). Data on the Web Management System: A Reference Model. *Digital Government Research: Governance in the Data Age*, p. 1-8.
- Oliveira, M. I., Lima, G. F., & Lóscio, B. F. (2019). Investigations into Data Ecosystems: a systematic mapping study. *Knowledge and Information Systems*, 61, p 589–630.
- Peppers, K.; Tuunanen, T.; Rothenberger, M.; Chatterjee, S. (2007). A design science research methodology for information systems research. *J. of Mgmt. Info. Systems*, 24 (3), p. 45–77.
- Penteadó, B. E. (2016). Correlational Analysis Between School Performance and Municipal Indicators in Brazil Supported by Linked Open Data. *WWW'16 Companion*, p. 507-512.
- Penteadó, B. E. (2017). Dados abertos educacionais: que informações temos disponíveis? *Congresso Brasileiro de Educação*, p. 507-512.
- Penteadó, B. E., Bittencourt, I. I.; Isotani, S. (2019a). Análise exploratória sobre a abertura de dados educacionais no Brasil: como torná-los prontos para o ecossistema da Web? *Revista Brasileira de Informática na Educação*, v.27 (1), p. 175-195.
- Penteadó, B. E., Bittencourt, I. I.; Isotani, S. (2019b). Modelo de referência para dados abertos educacionais em nível macro. *Simpósio Brasileiro de Informática na Educação*, p. 1808-1817.
- Penteadó, B. E., Bittencourt, I. I.; Isotani, S. (2019c). Metaprocesso para transformação de dados educacionais em dados conectados. *S. Brasileiro de Informática na Educação*, p. 1601-1610.
- Penteadó, B. E.; Maldonado, J. C.; Isotani, S. (2020). Process Model with Quality Control for the Production of High Quality Linked Open Gov. Data. *IEEE Latin America*, 19 (3), p 421-429.
- Penteadó, B. E.; Maldonado, J. C.; Isotani, S. (2021). Methodologies for publishing linked open government data on the web: a systematic mapping. *Semantic Web Journal* (em revisão).
- Penteadó, B. E.; Isotani, S. (2021). An analytics approach to teacher turnover. *Simpósio Brasileiro de Informática na Educação* [no prelo].
- Pereira, C. K., Siqueira, S. W. M., Nunes, B. P., Dietze, S. (2018). Linked Data in Education: A Survey and a Synthesis of Actual Research and Future Challenges. *IEEE Transactions on Learning Technologies*, vol. 11 (3), p. 400–412.
- Santos, O. A. R. (2014). Minha escola transparente: uma análise comparativa do uso de dados governamentais abertos na educação básica no Brasil e Inglaterra. *Dissertação (Mestrado prof. em Adm. Pública) – EBAP, FGV, Rio de Janeiro*.
- Santos, H.D.A.; Oliveira, M. I. S.; Lima, G. F. A. B.; Silva, K. M.; Muniz, R. I. V. C. S.; Lóscio, B. F. (2018). Investigations into data published and consumed on the web: a systematic mapping study. *Journal of the Brazilian Computer Society*, vol. 24, no. 14.
- Villazón-Terrazas, B.; Vilches-Blázquez, L.; Corcho, O; Gómez-Pérez, A. (2011). Methodological Guidelines for Publishing Government Linked Data. *Linking Government Data*, p. 27–49.
- W3C. Data on the Web Best Practices (2017). Disponível em: <https://www.w3.org/TR/dwbp>.