An architecture for monitoring public education policies based on big and open linked data

Bruno Elias Penteado[1], Seiji Isotani[2], Ig Ibert Bittencourt[2,3], Rafael Ferreira Mello[4], Ibsen Mateus Bittencourt[3]

[1] Instituto de Ciências Matemática e Computação – Universidade de São Paulo (USP)
13566-590 – São Carlos – SP – Brasil

[2] Harvard Graduate School of Education, Cambridge - MA, Estados Unidos

[3] Instituto de Computação – Universidade Federal de Alagoas (UFAL) – 57072-970 – Maceió – AL – Brasil

[4] Universidade Federal Rural de Pernambuco (UFRPE) – PE - Brasil

brunopenteado@alumni.usp.br, seiji_isotani@gse.harvard.edu, ig_bittencourt@gse.harvard.edu, rafael.mello@ufrpe.br, ibsen@feac.ufal.br

Abstract. Open government data (OGD) is increasingly being deployed in many countries. However, the role of OGD in policy-making monitoring has been explored only marginally due to its complexity and inter-organizational boundaries. Big Open Linked Data (BOLD) is a recent technological approach where data from multiple institutions with different data governance policies must be integrated to achieve a common goal. In this work, we present a BOLD infrastructure to support the monitoring of an educational public policy using a linked data approach. The design science research methodology (DSR-M) was followed in this study. We chose the PNE (National Education Plan) as a case study since it sets out goals for national education at different levels and perspectives for ten years. We used its goals and indicators to elicit requirements and evaluate the architecture. As a result, we created an infrastructure based on virtual knowledge graphs at a municipality level and used it to generate detailed and transparent statistics on calculating the goals. Finally, we highlight some implications and future challenges of our approach.

Resumo. Dados abertos governamentais (DAG) estão sendo cada vez mais implantados em muitos países. No entanto, o papel do DAG no monitoramento e na formulação de políticas foi explorado apenas marginalmente devido à sua complexidade e limites interorganizacionais. *Big Open Linked Data* (BOLD) é uma abordagem tecnológica recente em que dados de várias instituições com diferentes políticas de governança de dados devem ser integrados para alcançar um objetivo comum. Neste trabalho, apresentamos uma infraestrutura BOLD para apoiar o monitoramento de uma política pública educacional usando uma abordagem de dados vinculados. A metodologia *design science research* (DSR-M) foi seguida neste estudo. Escolhemos o PNE (Plano Nacional de Educação) como estudo de caso, pois ele traça metas para a educação

DOI: 10.5753/sbie.2022.225738

nacional em diferentes níveis e perspectivas para dez anos. Usamos seus objetivos e indicadores para elicitar requisitos e avaliar a arquitetura. Como resultado, criamos uma infraestrutura baseada em gráficos virtuais de conhecimento em nível de município e a usamos para gerar estatísticas detalhadas e transparentes sobre o cálculo das metas. Por fim, destacamos algumas implicações e desafios futuros de nossa abordagem.

1. Introduction

Governments worldwide have been continuously collecting data from multiple sectors for transparency, participation, efficiency, and accountability motivations [OGP, 2019] and applying efforts to make them public for scrutiny. Most governments are decentralized organizations in which every public body can determine their informational choices, usually under some form of central standardization or regulation, such as a minimum set of options for technologies or practices. One of the challenges of open government data (OGD) is to make this data reused by society as a form of social control, public policy monitoring, or economic development. Nevertheless, the publication of data grows exponentially, and there is a need to make them reach this potential impact, which can be achieved through semantic connections between sources, concepts, and data [Isotani and Bittencourt, 2015].

Big and open linked data (BOLD) is a recent emerging field in the technology-oriented business world [Janssen et al. 2015]. It refers to the integration of diverse data, without predefined restrictions or conditions of use, to create new insights [Janssen and Kuk 2016]. BOLD can be released by public and private organizations or individuals [Janssen et al. 2015] and can increase the reach of statistical and operational information and deepen the analysis of outcomes and impacts.

According to Janssen and Hoven (2015), Big and Open Linked Data significantly influences how governments operate and interact with the public. The BOLD acronym consists of the integration of three significant developments that affect our society. First, *open data* enables access to data without predefined restrictions or conditions of use. The opening of data can result in efficiency improvements and innovation, but also more transparency [Janssen et al., 2012]. Second, *linked data* is about connecting structured and machine-readable data that can be semantically queried [Bizer et al., 2009]. Third, *big data* is about large volumes of data from various sources that need to be processed [McAfee and Brynjolfsson, 2012]. The value of this data is created by combining data sources [Janssen et al., 2014], resulting in the availability of more data and the possibility of using business intelligence and data analytics to gain actionable insights from it [Holsapple et al., 2014]. Janssen et al. (2017) argue that big and open data should result in an open government instead of reinforcing current processes - i.e., part of a more extensive agenda that democratic countries set out to ensure values such as transparency, data availability, support civic engagement, deter corruption, and ensure accountability [OGP, 2019].

According to Janssen et al. (2017), data-driven innovation can result in a dramatic transformation of public sector systems and it can create societal benefits in multiple areas, like less pollution, fewer traffic jams, improved tracking of disease outbreaks, greater energy efficiency, new agriculture services, novel applications to transform citizen experience interacting online with government, and lower costs. Big and open data play a pivotal role in this transformation and collecting, combining, and sharing data from various sources has become an important means for public sector innovation. Although the potential applications of data analytics in many sectors have been investigated in many studies, in practice, these applications are rarely found in government processes or are, at most, only at the experimentation stage [Arnaboldi & Azzone, 2020]. Thus, the literature highlights the potential benefits of BOLD initiatives to fulfill the promises of open government data. However, they acknowledge that this is a very complex task, from different perspectives [Janssen et al., 2017]. Janssen and Hoven (2015) argue on the use of BOLD for public policies, suggesting the need for information architectures that protect privacy while maintaining transparency and suggesting the investigation on the impact of the use of BOLD on public organizations.

In this work, two research questions were developed: RQ1) *What is a feasible architecture for BOLD for a public education policy in the Brazilian government?* RQ2) *How can the data be modeled to meet the requirements of an actual public policy?* With these questions, we aimed to propose a feasible data architecture, in terms of performance and agility, and showed how it could be applied to an existing educational policy. Moreover, we show a data model that can be used to monitor indicators and metrics from public policies, illustrated with a case study from the Brazilian public education policy.

This paper is structured as follows: in section 2, we review the existing literature on the conceptualization of big, open, and linked data (BOLD) and the previous efforts made to understand this phenomenon and its impacts on organizations. In section 3, we specify the methodology applied to create the architecture and the data model to answer both research questions, detailing the case study adopted for validation. In section 4, we show the design choices, and resulting artifacts - the architecture and the data model developed to support the case study. Lastly, in section 5, we discuss the design process and highlight some implications for practitioners.

2. Related work

Much work with open government data was carried out in the last decade. National portals, such as from the USA (data.gov), the UK (data.gov.uk), and Brazil (dados.gov.br), were released in the late 2000s and early 2010s. Since then, thousands of datasets have been published from different public bodies, with their regulation and technological standards. Arnaboldi & Azzone (2020) argue that although data science projects have been claimed with potential innovations from the government's big data, they have still been explored marginally, demanding investigation because of their greater complexity and distinctive inter-organizational boundaries. Janssen et al. (2017) contend that many factors influence public sector innovation. Although this is an old challenge for public bodies, the current focus on data enables new types of applications.

Janssen et al. (2017) acknowledge the complexity of the BOLD approach and categorize these complexity factors in four interdependent dimensions: i) technical (technology readiness, systems, supporting infrastructure), ii) strategic and political (public values, societal problems addressed, data sharing licenses, culture, budget, incentives, etc.), iii) organizational (form of innovation, collaboration, cost-benefit analysis, risks, trust among stakeholders, capabilities of staff, project management, etc.), and iv) data governance (access to data, data quality, data reuse, data sharing, regulations, etc.). In this work, we address only the technical dimension, leaving room for future work in this matter.

Dwivedi et al. (2017) mapped the factors that most impacted the adoption of BOLD in organizations from a structural equation modeling with data gathered from a panel of specialists. From a total of 19 factors mapped from the specialists' experiences, three were considered the most fundamental: i) *technical infrastructure*: processing power, legacy systems, software access, storage capability, scalability and performance, and fragmentation, ii) *data quality*: completeness, accuracy, documentation, historical count, non-propriety, non-discriminatory, machine processable, interoperability, and iii) *external pressure*: market pressure, mimetic pressure, coercive pressure. In addition, they point out the main technical challenge of uncovering the effective data models and existing formalisms to handle data integration and transformations, as we partially map in our RQ2.

Several recent works have been developed in the Brazilian context over linked open government data. Ferreira et al. (2021) carried out systematic mapping of open education data in Brazil, pointing out ENEM and ENADE datasets as the ones most used for research on a national level, with few works proposing tools for the exploration. Penteado & Isotani (2017) elicited the educational datasets being deployed at the federal level in Brazil, categorizing them into four categories: census, academic, performance data, and policy expenditure. Penteado et al. (2019) analyzed the openness degree of datasets of each of these categories, showing a high degree of development for human consumption but not for automated processing, suggesting linked data as an option for this task. Some other works have explored BOLD for other applications of public interest, such as the relationship between education and public security (Auceli et al., 2020), chatbot for finding professional courses (Neto et al., 2020), metadata in OER repositories (De Paula et al., 2020), and correlations of school performance and socioeconomic factors (Penteado, 2016), but built from a perspective of answering specific questions with *ad-hoc* approaches. However, none of these works detailed a data model to represent educational indicators over the years and how to connect them with external sources. Besides, there is a lack for architecture designs to reuse current approaches for large legacy analytical databases.

We propose a BOLD architecture model that can help to address the current technical limitation listed above, supporting the monitoring of an educational Brazilian public policy with real-world data as a case study. This work builds upon these preview ones, proposing technological artifacts (architecture and data model) to combine these datasets and making them available as linked data for semantic querying, developed for public sector professionals with similar needs.

3. Methodology

The design science research method (DSR-M) proposed by Hevner et al. (2004) was followed in this study. DSR is a problem-solving paradigm that seeks to create artifacts through which information systems can be effectively and efficiently engineered. These artifacts are designed to interact with a problem in a context to improve something in that context [Wieringa, 2014]. It consists of a prescriptive solution paradigm where the organizational capabilities are extended together with the knowledge base through the creation and use of designed artifacts. The DSR method comprises two axes [March and

Smith, 1995]: the research *activities* and *products*. The research activities are composed of: problem definition (given in section 1), solution design and evaluation (section 4), and the extension of the body of knowledge about the problem according to the results (section 5). The research products can be: constructs, models, methods, or system instantiations. In this work, we propose an architectural and a data model as the resulting artifact, using the knowledge base from relational and multidimensional databases and linked data technologies and principles.

The illustrative case study. The PNE is an educational public policy composed of 20 goals in different levels of national education management, ranging from access to pre-schools to education funding, literacy, teachers' professional development, and so on. These 20 goals are further broken down into 47 indicators, with specific metrics for their calculation, as given by public technical notes. For the original monitoring of this policy, the estimates are made based on PNAD-C (National Household Sample Continuous Survey) - an annual survey carried out by IBGE (the Brazilian Institute of Geography and Statistics) with a high-level sampling for the entire country. The estimates are given only at the state and country levels. However, as much of the effort for carrying out this public policy is made at the municipality level, this project sought to extend this survey methodology to this higher level of granularity, meeting the needs of local education managers. To do that, we have been using open datasets from INEP (National Institute of Educational Research and Studies), MEC (Ministry of Education), and IBGE. The data has been available since 2014, the first year of this policy.

Evaluation. In order to evaluate our artifacts, we sought to model the PNE indicators as competency questions, a common approach in ontology evaluation methods. Here, we show the modeling for the goals in general but with results from Goal #1, Indicator B: "universalize, by 2016, early childhood education in preschool for children from 4 (four) to 5 (five) years old". As described earlier, this indicator is calculated only for the Brazilian states from PNAD-C data. Therefore, we calculate this metric from the basic education census, with the total enrolments by the municipality, and we estimate the population of children aged 4 or 5 for a given year for every municipality in the country.

4. Artifact description and evaluation

In this section we detail the design search process and the resulting artifacts, as described in Gregor & Hevner (2013) to describe the artifact design. For the first research question, two steps were taken. Firstly, we applied database analytics methods such as data warehouse modeling with a star schema, with the goals' indicators as fact tables and the related information (municipalities, states, indicators descriptions) as dimensions. Then, as a preprocessing step, the original files, with different formats and granularities, were ingested into the schema for normalization. Although modern approaches, such as data lakes or NoSQL databases, could be applied over the original files, we considered the mature infrastructure of data warehouses to handle large volumes of data, observing some of the fundamental requirements for BOLD, as given by Dwivedi et al. (2017) - in particular, the technical infrastructure. Besides, for agile prototyping and since we had no prior applications specified to consume these data, we aimed to make it simple and to further expand it accordingly.

Next, a design choice was made for transforming the existing data into its semantic representation. Two main approaches are available in the literature [Hartig & Langegger, 2010]: i) the materialized view, via triplification of the underlying relational data and ii) virtual knowledge graphs (VKG), which maps relational databases to its semantic representation. We employed the VKG approach, which projects the original data into the semantic space using mapping files, specifying the semantic annotations over the original attributes and classes, in a sense established by the W3C Direct Mapping Recommendation [W3C, 2012] for generating linked data from relational data. In this approach, the data is not replicated and materialized into RDF triples; instead, a layer is added on top of the existing relational database. In this layer, the relational data is mapped to the RDF format using semantic vocabularies - preferably, the ones already well established by the community, such as FOAF[1], Schema.org[2], or Dublin Core[3]. On the other hand, the alternative materialization approach aims to replicate all the original data into a triplestore in the RDF format, resulting in a much higher requirement for storage. For a comprehensive comparison of the advantages of each approach, please refer to Hartig & Langegger (2010).

Figure 1 illustrates the generic architecture of the virtual knowledge graph over the transformed data warehouse. The bottom of the architecture shows the relational data warehouse, built from different datasets to calculate the indicators' values. On top of it, we applied the VKG, creating the mappings for each attribute of interest ("virtual knowledge graph" layer). We used D2RQ Server (http://d2rq.org), a system for accessing relational databases as virtual, read-only RDF graphs. It offers RDF-based access to the content of relational databases without having to replicate it into an RDF store. This software exposes the data mapped semantically in different access points: queries over a SPARQL endpoint using the respective language, dereferenciation of the resources using a linked data interface, and a Web API to query the data using the fundamentals of the Web ("data access" layer). Thus, this data can be flexibly queried by software agents in an automated manner or by humans exploring or integrating manually these data (the "consumers" layer).

[1] https://xmlns.com/foaf/spec/
[2] https://schema.org/
[3] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/

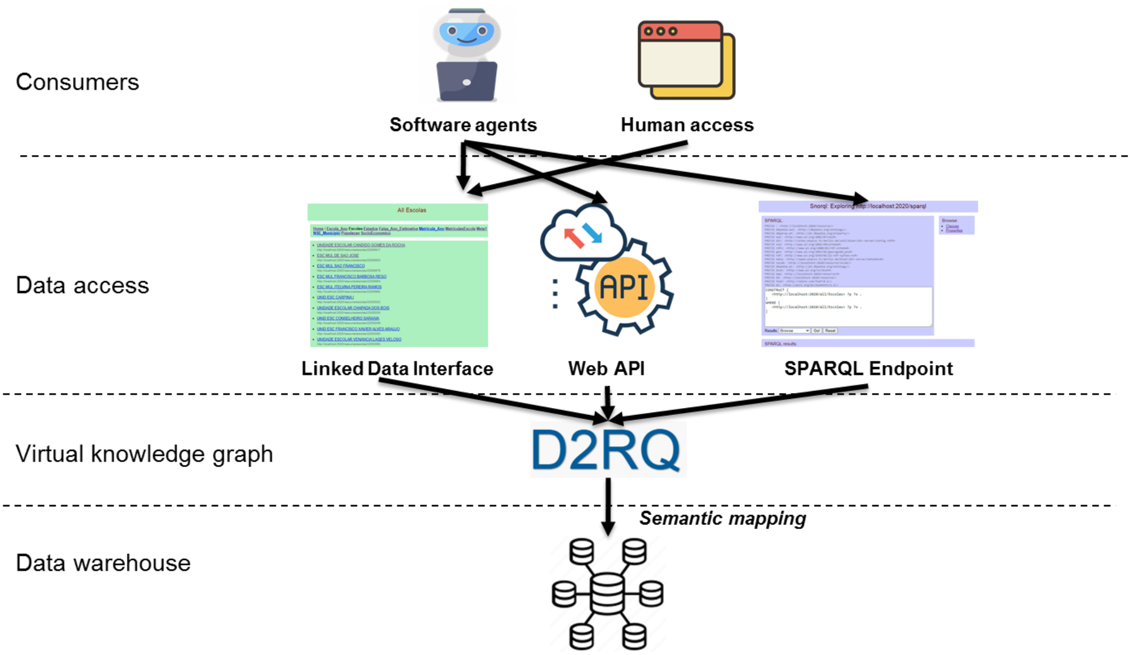

Figure 1. The architectural model proposed for BOLD to handle the monitoring of an educational public policy in Brazil.

For our second research question, regarding the semantic modeling of the educational policy data, we have taken three steps. First, we mapped educational vocabularies suited for this task. However, very few vocabularies were developed to address educational administrative data. The ontologies developed by Penteado (2016) and Penteado et al. (2019b) was used for modeling the educational indicators. The other terms had to be created from scratch using Protégé and deployed on the Web. The upper classes (School, Municipality, State) were aligned to classes of Schema.org. As an additional interlinking step at the instance level, all the municipalities were mapped to their respective resources in DBpedia (https://dbpedia.org). Although this was not a requirement, it is a best practice to integrate new data into the linked open data cloud (https://lod-cloud.net).

Second, we had to model the data in an optimized manner. As a longitudinal monitoring, we have the same entity with different measurements over the years. In the data warehouse, it is not in a normalized form, but it becomes easier to understand and query when present in a normal form. To that, we used blank nodes. Although it is not a functional requirement, and it is a best practice to avoid this, it is often the case where an intermediate data structure should be created to optimize the comprehensibility and lower the complexity of querying the data. Furthermore, the query results are transferred through the Web in text format - thus, the optimization of the query results also brings technical advantages. Figure 2 shows the modeling of Goal #1 ("Meta"), indicator B (universalization of pre-schooling) for a particular municipality ("Alta Floresta d'Oeste") in two consecutive years (2014 and 2015) with their respective values, according to these modeling choices.

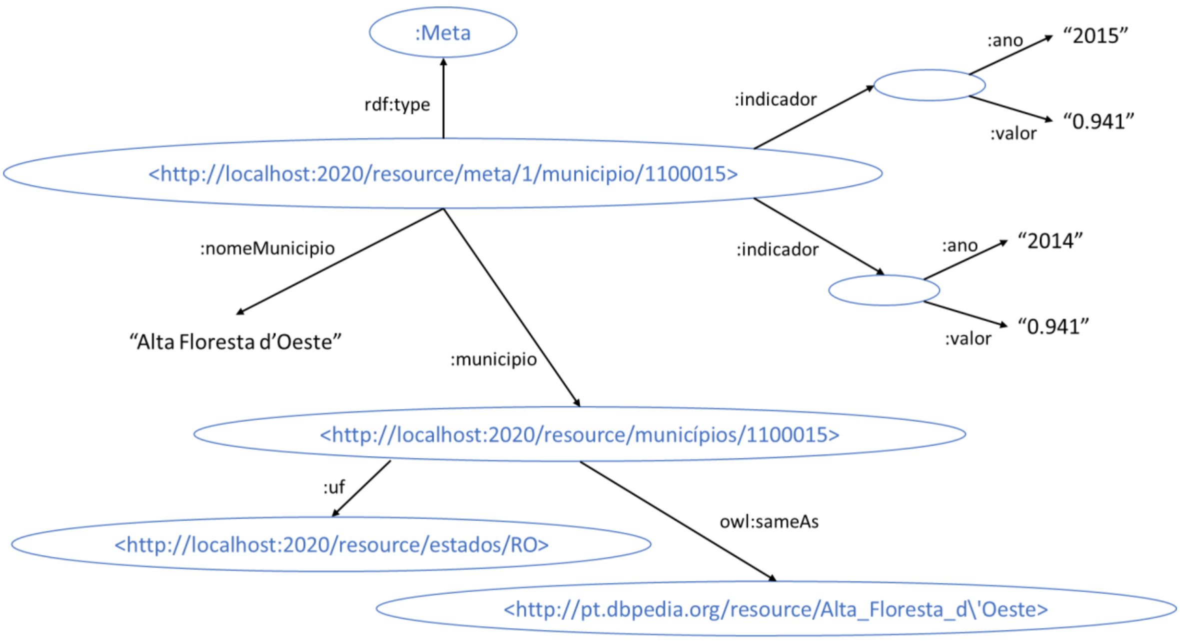

Figure 2. The RDF data model of the instantiation of the indicator metrics (Goal #1, indicator B) values for a municipality (Alta Floresta d’Oeste - RO) in 2014 and 2015.

Finally, we prepared the metadata of these data, at the service-, catalog- and resource levels, providing information such as data sharing licenses, data provenance, and data versioning, as recommended by Penteado (2020). As pointed out by the W3C (2017), these features increase the chance of reuse over this data. In addition, the metadata is considered part of the knowledge graph, being available for querying by software agents or humans.

Figure 2 also illustrates the data model obtained in response to a query for the municipality with code 1100015 (“Alta Floresta d’Oeste”) in Goal 1, represented by the URI <http://localhost:2020/resource/meta/1/municipio/1100015> dereferenced by the data access layer. The data structure returned contains all related attributes: type (rdf:type predicate), label (:nomeMunicipio), the indicators (:indicators), and the resource for the municipality’s entity. This last one could also be dereferenced to obtain more information about the municipality, such as the corresponding name, state of the country, and the DBpedia resource, with even further information.

5. Discussion

Governments worldwide are trying to handle the big data they continuously collect in different areas to extract value in their relationship with citizens. In this paper, we applied an illustrative case study investigating a public education policy, PNE, which sets out 20 goals for the years 2014-2024 in Brazil. This approach extends the official measurement from INEP, which collects annual sampled data based on states’ data. Here, we use different public databases and methods to predict these data at the municipality level - aiming to help district administrators pursue these goals locally. The choice of this illustrative case study helped to understand some of the complexities cited in the literature. Multiple datasets from different agencies, created to be independent, were set out to be combined to extract value for a particular application. Some measurements had to be estimated since the official methodology did not meet the requirements of this solution, such as the population census performed every ten years. Its longitudinal nature made clear the need for the data model to handle this temporal aspect. The large volume of data over the years also impacted the design choices for the artifacts, making them as compact as possible during the semantic modeling.

The prescriptive aspect of DSR made it possible to use existing knowledge bases to address the problems posed in our research questions and create artifacts that can be replicated for similar classes of problems - the monitoring of public policies from open datasets. Storage considerations and agility in prototyping were essential to the choice of virtual knowledge graphs, which allowed the semantic data to be accessed by multiple points. It also presented the advantage of not changing the underlying database. Being in linked open data format, this solution improves the transparency of the policy since all the indicators' measurements can be traced semantically, facilitating replication studies.

The three BOLD axes were employed in this study. *Open data*, by collecting open datasets from Brazilian federal agencies and making them accessible; *big data*, with the volume, veracity, and variety of public datasets; and *linked data*, integrating and connecting structured and machine-readable data that can be semantically queried using Web standards. Innovation in government is about finding new ways to improve society, the government itself, and the relationship between the government and the public [Janssen et al., 2017].

In future work, we aim to expand this methodology to the other goals of PNE and improve the architecture to meet the standards of W3C's recommendations on data sharing on the Web [W3C, 2017]. Furthermore, this solution also opens room for the development of semantic applications that can automatically query and make inferences on the data, as proposed by Berners-Lee's vision of the Semantic Web [Berners-Lee et al., 2001].

6. Conclusion

In this work we presented an architecture to monitor educational policies based on big and open linked data technologies. This is a common scenario where a complex public policy requires the combination of different data sources from different agencies to calculate managerial indicators. Through the selected case study used here we argue that the method and the architecture developed here generalize to similar contexts, contributing to the scientific literature of BOLD artifacts, electronic government, and information management in educational settings. On the other hand, we emphasize the practical implication of the artifacts, providing basis for the combination of different sources and making it transparently available for queries from human users and software agents.

References

Arnaboldi, M., Azzone, G. (2020). Data science in the design of public policies: dispelling the obscurity in matching policy demand and data offer. Heliyon, vol. 6 (6), e04300. DOI: 10.1016/j.heliyon.2020.e04300.

Auceli, P. H. S., Belizario, M. G., Kozievitch, N. P., Berardi, R. C. G. (2020). Uma discussão sobre a relação entre segurança e desempenho acadêmico de estudantes do ensino médio apoiada por dados abertos conectados de Curitiba. Anais do XXXI

Simpósio Brasileiro de Informática na Educação (SBIE 2020), p. 1002-1011. DOI: 10.5753/cbie.sbie.2020.1002.

Berners-Lee, T.; Hendler, J.; Lassila, O. The SemanticWeb: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. 2001.

Bizer, C., Heath, T., Berners-Lee, T. (2009). Linked data — The story so far. International Journal on Semantic Web, 5 (3) (2009), p. 1-22.

De Paula, L. B., Deus, W. S., Barbosa, E. F. (2020). Análise de Repositórios de REAs em Relação ao Uso dos Padrões de Linked Open Data. Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020), p. 312-321. DOI: 10.5753/cbie.sbie.2020.312.

Dwivedi, Y. K., Janssen, M., Slade, E.L., Rana, N. P., Weerakkody, V., Millard, J., Hidders, J., Snijders, D. (2017). Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural modelling. Information Systems Frontiers, vol. 19, p. 197–212. DOI: 10.1007/s10796-016-9675-5.

Ferreira, L. A., Rodrigues, R. L., Souza, R. N. P. M. (2021). Dados Abertos Educacionais Brasileiros: Um Mapeamento Sistemático da Literatura. Anais do XXXII Simpósio Brasileiro de Informática na Educação (SBIE 2021), p. 1186-1195. DOI: 10.5753/sbie.2021.218158.

Gregor, S., Hevner, A. (2013). Positioning and presenting Design Science Research for maximum impact. MIS Quarterly, 37 (2), p. 337-355. DOI: 10.25300/MISQ/2013/37.2.01.

Hartig, O. & Langegger, A. (2010). A database perspective on consuming linked data on the web. Datenbank-Spektrum, 10:57–66. DOI: 10.1007/s13222-010-0021-7.

Holsapple, C., Lee-Post, A., Pakath R. (2014). A unified foundation for business analytics Decision Support Systems, 64, p. 130-141, 10.1016/j.dss.2014.05.013.

Isotani, S., Bittencourt, I. I. (2015). Dados abertos conectados. Novatec.

Janssen, M., Charalabidis, Y., Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information Systems Management, 29 (4) (2012), pp. 258-268, 10.1080/10580530.2012.716740.

Janssen, M., Estevez, E., Janowski, T. (2014). Interoperability in big, open, and linked data — Organizational maturity, capabilities, and data portfolios. Computer, 47 (10) (2014), pp. 44-49, 10.1109/MC.2014.290.

Janssen, M., Hoven, J. v. d. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? Government Information Quarterly Volume 32, Issue 4, October 2015, Pages 363-368.

Janssen, M., Kuk, G. (2016) Big and Open Linked Data (BOLD) in research, policy, and practice, Journal of Organizational Computing and Electronic Commerce, 26:1-2, p. 3-13, DOI: 10.1080/10919392.2015.1124005.

Janssen, M., Konopnicki, D., Snowdon, J.L., Ojo, A. (2017). Driving public sector innovation using big and open linked data (BOLD). Information Systems Frontiers, 19, p .189–195. DOI: 10.1007/s10796-017-9746-2.

Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design science in information systems research. MIS Q. 2004, 28, 75–105.

McAfee, A., Brynjolfsson, E. (2012). Big data: The management revolution. Harvard Business Review (90), p. 60-66.

March, S. T., Smith, G. F. (1995). Design and natural science research on information technology. Decision Support Systems, v. 15, n. 4, p. 251 – 266, 1995. ISSN 0167-9236.

Neto, A. J. M., Silva, C. E., Anjos, W. F., Dorça, F. A. (2020). Uma Abordagem Baseada em Dados Abertos Conectados e Chatbot para Disponibilizar o Catálogo de Cursos da Rede Federal de Educação Profissional, Científica e Tecnológica. Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020), p. 1263-1272. DOI: 10.5753/cbie.sbie.2020.1263.

OGP. Open Government Partnership - Open Government Declaration. (2019). Disponível em: https://www.opengovpartnership.org/open-government-declaration. Acesso em: 17/07/2022.

Penteado, B. E. (2020). Modelo de infraestrutura para publicação de dados abertos governamentais conectados de qualidade. Tese de Doutorado, Instituo de Ciências Matemáticas e da Computação, Universidade de São Paulo, 296 p. DOI: 10.11606/T.55.2020.tde-14092020-175138.

Penteado, B. E., Bittencourt, I. I., Isotani, S. (2019). 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, n. 01, p. 175.

Penteado, B. E., Bittencourt, I. I., Isotani, S. (2019b). Modelo de referência para dados abertos educacionais em nível macro. XXX Simpósio Brasileiro de Informática na Educação. DOI: 10.5753/cbie.sbie.2019.1808.

Penteado, B. E., Isotani, S. (2017). Dados abertos educacionais: que informações temos disponíveis? VI Congresso Brasileiro de Educação, vol. 4, p. 1933-1938.

Penteado, B. E. (2016). Correlational Analysis Between School Performance and Municipal Indicators in Brazil Supported by Linked Open Data. Proceedings of the 25th Intl. Conf. Companion on World Wide Web - WWW '16 Companion.

W3C. (2012). A Direct Mapping of Relational Data to RDF. https://www.w3.org/TR/rdb-direct-mapping.

W3C. (2017). Data on the Web Best Practices. https://www.w3.org/TR/dwbp/.

Wieringa, R.J. Design Science Methodology for Information Systems and Software Engineering; Springer: Berlin, Germany, 2014