

Modelo de referência para dados abertos educacionais em nível macro

Bruno Elias Penteadó¹, Ig Ibert Bittencourt², Seiji Isotani¹

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

²Instituto de Computação – Universidade Federal de Alagoas (UFAL) – 57072-970 –
Maceió – AL - Brasil

brunopenteadó@usp.br, ig.ibert@ic.ufal.br, sisotani@icmc.usp.br

Abstract. *Open data has been published by governments worldwide in the last years. The educational domains in often one of the priority ones, presenting multiple perspectives represented by these datasets, which also brings great complexity for its integration and reuse. In this work, we present a domain ontology that enables to describe education data in a macro-level, derived from open government data. As result, it is possible to cover a big part of the datasets officially available and that can be reused in data integration among different information systems. This work contributes with a domain ontology for education open government data and a method to derive this ontology from data.*

Resumo. *Dados abertos têm sido publicados por governos do mundo todo nos últimos anos. O domínio educacional é um dos prioritários, contendo diversas perspectivas representadas por esses dados, o que traz também grande complexidade para sua integração e seu reuso. Neste trabalho apresentamos uma ontologia de domínio que possibilita descrever dados educacionais em nível macro, a partir dos dados abertos governamentais. Como resultado, é possível obter uma cobertura de quase todos os conjuntos de dados disponíveis oficialmente e que pode ser usado na integração de dados entre diferentes sistemas de informação. Este artigo propõe como contribuição uma ontologia de domínio para dados abertos governamentais educacionais e um método de derivar tal ontologia a partir de dados já existentes.*

1. Introdução

Dados abertos são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa - sujeitos, no máximo, à exigência de atribuição à fonte original e compartilhamento pelas mesmas licenças em que as informações foram apresentadas¹. Desde o final da década passada, governos do mundo todo têm destinado esforços para publicar seus dados, coletados junto à sociedade, concebendo-os como propriedade pública e desenvolvendo leis e portais para regular sua coleta, processamento

¹ <https://opendefinition.org/>

e distribuição. E um dos desafios dos dados abertos governamentais está em fazer com que esses dados sejam reutilizados pela sociedade, como forma de controle social, monitoramento de políticas públicas ou mesmo o desenvolvimento econômico, por meio de novos produtos e serviços [Janssen et al. 2012]. No entanto, a geração de dados cresce em ritmo exponencial, e existe a necessidade aproveitá-los por meio da análise das conexões semânticas entre conceitos e relações esses dados [Isotani e Bittencourt, 2015].

No domínio educacional, Santos (2014) classifica a geração e uso dos dados em 3 níveis: *macro* (dados em nível nacional ou regional), *meso* (nível escolar) e *micro* (nível de sala de aula ou aluno). Os dados de nível macro têm a função de fornecer uma visão holística sobre os resultados de políticas públicas ao longo do tempo, como por exemplo por meio dos censos escolares de diversos níveis como pelos gastos em determinada política pública. Assim, os dados abertos neste nível têm um público-alvo diversificado, atendendo a necessidades de órgãos públicos, pesquisadores de diferentes áreas e empresas que desenvolvem serviços que exploram esse conhecimento armazenado.

Essa grande disponibilidade de dados apresenta também uma complexidade de integrar as informações, já que estão disponíveis em diferentes formatos e descritos de maneiras diferentes dentro dos repositórios oficiais, exigindo grande conhecimento implícito de domínio para seu processamento. A não integração desses dados traz dificuldades para os consumidores dos dados como: dificuldade em conectar diferentes fontes de dados (dentro e fora do escopo educacional), já que identificadores não são necessariamente padronizados e existem diferentes graus de granularidade de informação; a existência de diferentes nomes e descrições para a mesma informação; dificuldade na criação de aplicações que possam inferir informações a partir dos relacionamentos semânticos estabelecidos. Neste trabalho, estruturamos tais informações em torno de um pequeno conjunto de entidades relacionadas aos contextos educacionais representados por esses conjuntos de dados. Para este fim, propomos um modelo de referência para esses dados educacionais em nível macro, facilitando a anotação semântica dos conjuntos de dados abertos disponíveis e habilitando a exploração inequívoca dos relacionamentos entre as diferentes fontes de dados.

Modelos de referência são arcabouços abstratos para compreender relacionamentos entre entidades em determinado contexto e para o desenvolvimento de padrões ou especificações consistentes que dão suporte a este contexto, de modo a encorajar uma comunicação clara destes conceitos, usados para demonstrar ou explicar um determinado domínio a usuários não especialistas [OASIS, 2014]. Uma possível representação desses modelos de dados é feita por meio de ontologias - artefatos formais de especificação de termos em um domínio e de suas relações entre si [Gruber, 1993].

Diferentes taxonomias de ontologia podem ser encontradas na literatura [Guarino, 1997]. Para nosso contexto, destacamos uma diferenciação comum: ontologias de *domínio* e de *tarefa*. O objetivo de uma ontologia de *domínio* é o de permitir a troca de informações e integração entre diferentes fontes de dados, ao fornecer definições semânticas e desambiguações para transformar fontes de informação locais em recursos

globais, seja em contexto local ou espalhado pela Web. Deste modo, esta ontologia de domínio não objetiva propor o que macrodados educacionais devem documentar, mas explicam a lógica do que atualmente documentam e, assim, permite a sua interoperabilidade semântica. Já as ontologias de *tarefa* representam processos e atividades para resolver um determinado problema abstraído o contexto do domínio. Este trabalho propõe uma ontologia de domínio para dados educacionais em nível macro, tomando como ponto de partida os dados educacionais já publicados e amadurecidos ao longo das últimas décadas.

2. Trabalhos relacionados

Diversas ontologias no domínio educacional foram desenvolvidas para diferentes domínios e níveis de granularidade, porém com maior ênfase no nível micro, mais ligado aos processos mais diretamente ligados à aprendizagem dos alunos. Por exemplo, no nível micro, Carchedi et al. (2018) compilaram diferentes *frameworks* com informações para aplicações de *learning analytics*, como perfil de aluno, envios de mensagens, material acessado em um ambiente virtual de aprendizagem, tempo de uso da ferramenta, resultados de testes, entre outros dados. Outros trabalhos também desenvolveram ontologias para descrever materiais instrucionais (recursos educacionais abertos), como os padrões IEEE-LOM, Dublin Core e OBAA (Gluz & Vicari, 2011).

Neste trabalho, adotamos o nível de granularidade *macro*, que trata dos dados gerados por instâncias superiores nos órgãos públicos brasileiros. Além do mais, tais dados são usados para compreender a evolução do aprendizado no país, monitorar os gastos e efetividade de políticas públicas, dentre outros usos. Poucos trabalhos abordaram a modelagem de macrodados educacionais na literatura. Penteado (2016) apresenta uma ontologia de domínio para tratar de indicadores de diferentes naturezas - educacionais, econômicas e sociais - para analisar a correlação do desempenho acadêmico em função dessas outras variáveis socioeconômicas, em nível municipal. Cabral et al. (2012) modelaram microdados do ENEM usando duas ontologias: *enem* e *geopoliticabr*, que descrevem respectivamente, competências e notas no ENEM e o nome dos municípios e sua geolocalização. Ambos os trabalhos citados se concentraram no desenvolvimento de ontologias de aplicação, ou seja, com escopo mais reduzido e finalidade específica. Este artigo visa desenvolver uma ontologia de domínio, que seja generalizável entre diferentes aplicações e que, ainda assim, reflita com grande cobertura os dados já disponibilizados.

3. Procedimento metodológico

Adotamos neste trabalho a metodologia de pesquisa qualitativa de *design science research* (DSR). Ela foi selecionada por oferecer *framework* metodológico para a criação e avaliação de artefatos tecnológicos. Trata-se de um paradigma de soluções de problemas, prescritivo em vez de explicativo, nas quais os limites das capacidades organizacionais são estendidos junto com o conhecimento do domínio do problema por meio da construção e aplicação de artefatos projetados [Hevner, 2010]. A DSR é composta por dois eixos [March & Smith, 1995]: as *atividades* de pesquisa e os *produtos*

da pesquisa. As atividades de pesquisa compreendem os passos necessários para definir o problema (definido na introdução), projetar uma solução (seção 4), avaliá-la (seção 5) e aumentar a base de conhecimento sobre o problema baseado nos resultados (seção 6). Como produto da pesquisa, o artefato pode assumir alguma das seguintes formas: construtos, modelos, métodos e implementações de sistemas.

Adotamos uma abordagem *bottom-up* para a construção da ontologia de domínio, em que o mapeamento de entidades e relacionamentos são criados a partir de dados (instâncias) já existentes, generalizando os conceitos a partir dos dados individuais. Essa escolha foi feita de modo a reaproveitar a grande quantidade de dados abertos já publicados no domínio educacional e que se apresentam de modo estável ao longo dos anos. Uma suposição adotada é a de que tais dados já embutem uma estrutura definida e apurada durante a modelagem dos dados presentes na fonte.

Para a construção do modelo de referência do domínio, foram investigados todos os conjuntos de dados presentes no portal de dados abertos do governo brasileiro (dados.gov.br) no âmbito das organizações que gerenciam os aspectos macroeducacionais no Brasil (MEC e INEP), como a formulação e monitoramento de políticas públicas, censos de instituições e atores educacionais, dados acadêmicos e indicadores de desempenho (total de 64 conjuntos de dados e 112 distribuições). Foram excluídos conjuntos de dados pontuais, que foram publicados apenas uma vez, já que sugere que essas informações podem não ser tão úteis e dificilmente serão republicados [Penteadó & Isotani, 2017]. Cada conjunto de dados foi classificado conforme as categorias estabelecidas por Santos (2014): dados censitários, dados orçamentários, dados acadêmicos e indicadores de desempenho. Em seguida, foram extraídas as entidades principais que descrevem esses conjuntos, suas particularidades e como elas se relacionam entre si, materializando-se no propósito de cada conjunto de dados. Além disso, a ontologia proposta por Penteadó (2016) foi reutilizada, uma vez que apresenta uma parte dos mesmos conceitos abstratos que descrevem esse domínio. A Figura 1 ilustra este processo.

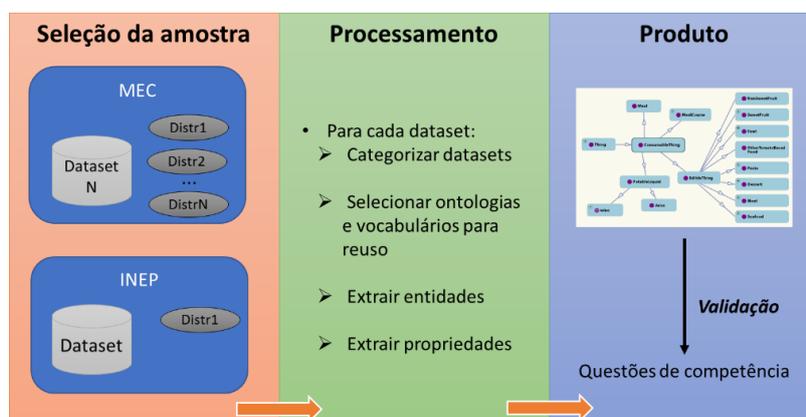


Figura 1. Procedimento metodológico adotado neste trabalho.

Utilizamos a metodologia *Ontology 101* [Noy & McGuinness, 2001] para extrair uma representação ontológica a partir dos dados já publicados, com os seguintes passos:

1. *Determinar o domínio e escopo da ontologia*: o escopo desta ontologia deve cobrir informações governamentais sobre o domínio educacional; deverá ser usada para fornecer informações sobre indicadores socioeconômicos em diferentes níveis de granularidade (desde escola até federal); poderá ser usada por diferentes atores para explorar dados já publicados. Definimos as seguintes questões de competência [Gruninger & Fox, 1995] para fins de validação da ontologia, cobrindo cada categoria de fonte de dados com uma questão:
 - a. *Orçamentário*: Quanto o município de São Paulo recebeu de verbas do programa PDDE em 2014?
 - b. *Censitário*: Qual a porcentagem de escolas municipais no Brasil que têm banda larga, em 2018?
 - c. *Acadêmico*: Qual porcentagem de escolas no Brasil apresenta níveis socioeconômicos altos ou muito altos (grupos 6 e 7), em 2017?
 - d. *Desempenho*: Qual a nota IDEB em Matemática para as séries finais do Ensino Fundamental, das escolas municipais de São Paulo em 2011?
2. *Considerar o reuso de ontologias*: usamos a ontologia Dbpedia² como fonte de dados de definição de classes a ser estendida para nosso domínio, pois se trata de uma ontologia extensa e com mapeamento de equivalência entre classes de diferentes fontes, como Schema.org³ e Wikidata⁴. Outros vocabulários foram adotados para as propriedades como Dublin Core⁵ (que descreve artefatos digitais), Time Ontology⁶ (para descrever intervalos de tempo), Organization Ontology⁷ (descrever estruturas organizacionais).
3. *Enumerar termos importantes na ontologia, definir as classes e sua hierarquia e definir as propriedades das classes*: esses passos são muito relacionados entre si, não ocorrendo de forma sequencial, mas de forma incremental e cíclica [Noy & McGuinness, 2001]. Como termos importantes, consideramos a tipologia de definida por Santos (2014). Além disso, consideramos cada dataset como uma nova classe, subclasse de uma das categorias e seus campos como propriedades da respectiva classe, similar ao proposto por Berners-Lee (2001) para mapeamento de dados relacionais para a Web. E, para cada dataset, identificamos quais são as propriedades mais importantes para o escopo deste artigo - uma vez que um *dataset* censitário, por exemplo, contém centenas de propriedades - e quais as relações dos *datasets* com as classes principais.
4. *Definir as facetas das propriedades*: neste trabalho não foram mapeadas as restrições, o que será feito em futuras iterações de refinamento deste modelo.

² <http://dbpedia.org/ontology>

³ <https://schema.org/docs/schemas.html>

⁴ <https://www.wikidata.org>

⁵ <http://dublincore.org/schemas/rdfs/>

⁶ <https://www.w3.org/TR/owl-time/>

⁷ <https://www.w3.org/TR/vocab-org/>

5. *Criar instâncias*: no escopo deste trabalho, apenas os dados referentes a um município foram compilados, dada a grande quantidade de recursos necessários para processar todos os conjuntos de dados disponíveis.

4. Resultado

O procedimento metodológico citado na seção 3 permitiu derivar um conjunto relativamente pequeno de classes e propriedades⁸. Três agrupamentos conceituais emergiram: *i*) modelo de dados de um indicador, *ii*) os diferentes tipos de indicadores com suas características particulares e *iii*) o alvo do indicador, ou seja, a quem se aplica aquilo que foi medido. As Figuras 2, 3 e 4, respectivamente, mostram diferentes visões da ontologia resultante.

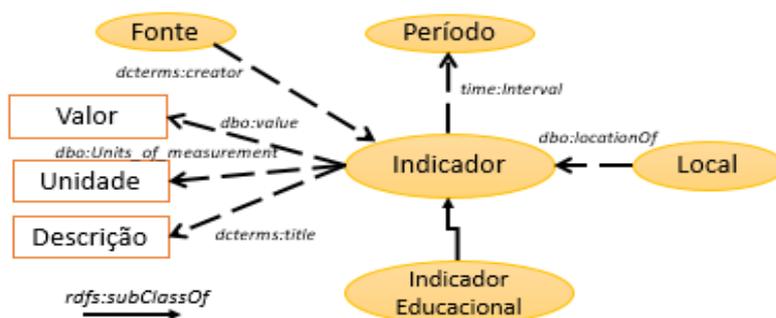


Figura 2. Estrutura básica de dados de um indicador.

Na Figura 2 são exibidos os principais componentes de um indicador - neste caso, um indicador educacional. O indicador (descrição), criado por um órgão (Fonte), aponta qual agente (Local) demonstrou qual valor (valor e unidade) em uma determinada fatia de tempo (Período).

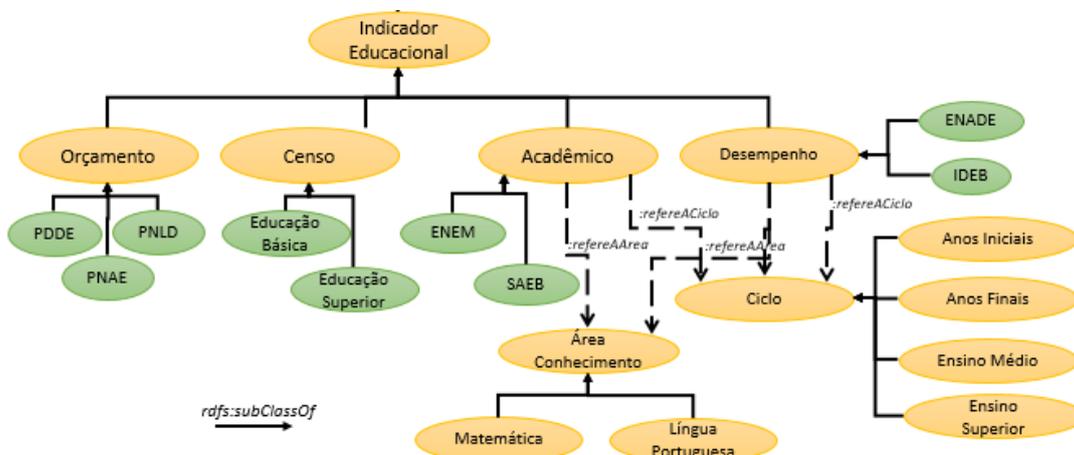


Figura 3. Características dos diferentes indicadores educacionais. Em verde, exemplos concretos de *datasets* existentes.

A Figura 3 detalha as particularidades dos indicadores educacionais disponíveis como dados abertos. A lista de classes não é exaustiva, por razão de espaço neste artigo.

⁸ Disponível em: https://github.com/brunopenteado/sbie2019/blob/master/Macroeducacao_v2.owl

Os indicadores educacionais são categorizados conforme explicitado anteriormente em 4 categorias: i) orçamento: lista programas educacionais que liberaram seus dados e que tem como peculiaridade que a unidade dos valores dos indicadores são em moeda (reais); ii) censitários: dados que configuram a existência de entidades (escolas, alunos, professores, vagas) no cenário educacional; iii) acadêmico e de desempenho: ambas trazem como contextualização: a) o ciclo a que se destina: ensino fundamental (anos iniciais e finais), ensino médio ou ensino superior; b) a área de conhecimento envolvida: as mais comuns, Língua Portuguesa e Matemática, mas no entanto varia a cada indicador, como por exemplo o ENEM ou o ENADE. Os indicadores de desempenho se diferenciam por trazerem um cálculo que agrega diferentes medidas acadêmicas em um único fator numérico.

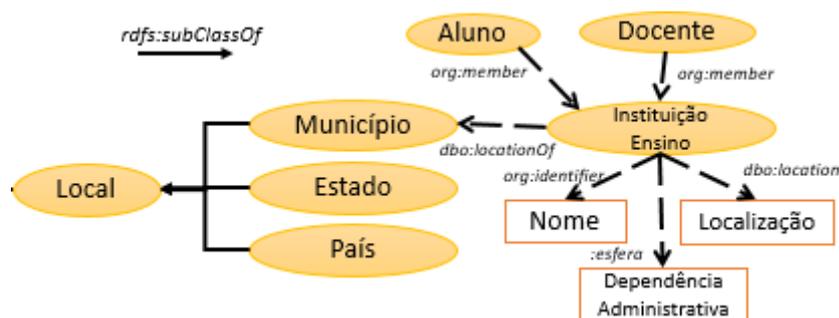


Figura 4. Alvo dos indicadores educacionais.

A Figura 4 demonstra quais foram os principais agentes modelados pelos dados abertos. Muitos dos indicadores, de todas as categorias, trazem informações sobre uma região administrativa (Local), seja ela um município, o estado que o contém ou o próprio nível nacional. Além disso, alguns conjuntos de dados trazem informações sobre as instituições de ensino (básico ou superior), que podem chegar ao nível das pessoas que as compõem (Aluno ou Docente). Em geral, os seguintes valores são mais reutilizados: nome da instituição, sua dependência administrativa (privada, municipal, estadual, federal) e sua localização (urbana ou rural).

A Figura 5 mostra as estruturas de dados da ontologia, contendo algumas das instâncias na base, para efeitos de ilustração da implementação e respondendo às questões de competência *a* e *d* definidas na seção 3. Os dados brutos foram mapeados via script para gerarem a sintaxe Turtle a ser importada na ontologia.

5. Avaliação

Para avaliar a expressividade da ontologia proposta, modelamos diferentes conjuntos de dados, mais precisamente um conjunto de dados por categoria, já que *datasets* da mesma categoria tendem a compartilhar muito de suas estruturas entre si. Para isso descrevemos as questões de competência propostas na seção 3 e, com base na ontologia extraída dos dados em formato bruto, exemplificada na Figura 5, conseguimos responder a cada uma das questões de competência propostas, conforme indicado na Tabela 1.

```

@prefix : <http://sbie2019.org/ns#> .
@prefix time: <https://www.w3.org/TR/owl-time/#time:> .
@prefix dbo: <https://dbpedia.org/ontology/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

# definição dos períodos desejados
:p2014 a :Período .
:p2014 time:years "2014"^^xsd:decimal .
:p2011 a :Período .
:p2011 time:years "2011"^^xsd:decimal .

# Indicador PDDE
:PDDE rdfs:subClassOf :Orçamento .
:pdde14 a :PDDE .
:pdde14 time:interval :p2014 .
:pdde14 dcterms:creator <http://www.inep.gov.br/> .
:pdde14 dbo:value "57643863,23"^^xsd:decimal .
:pdde14 dbo:Unity_of_measurement "reais (R$)" .
:pdde14 dcterms:title "PDDE 2014" .
:Sao_Paulo :hasIndicator :pdde14 .

# Indicador IDEB
:IDEB rdfs:subClassOf :Desempenho .
:ideb2011 a :IDEB .
:ideb2011 time:interval :p2011 .
:ideb2011 :refersToCycle :AnosFinais .
:ideb2011 :refersToArea :Matematica .
:ideb2011 dcterms:creator <http://www.inep.gov.br/> .
:ideb2011 dbo:value "4.2"^^xsd:decimal .
:ideb2011 dbo:Unity_of_measurement "" .
:Sao_Paulo :hasIndicator :ideb2011 .

# Schema básico
:Indicador rdfs:subClassOf owl:Class .
:Local rdfs:subClassOf owl:Class .
:Indicadoreducacional rdfs:subClassOf :Indicador .
:Orçamento rdfs:subClassOf :Indicadoreducacional .
:Desempenho rdfs:subClassOf :Indicadoreducacional .
:Município rdfs:subClassOf :Local .
:Período rdfs:subClassOf owl:Class .
:Ciclo rdfs:subClassOf owl:Class .
:AnosFinais rdfs:subClassOf :Ciclo .
:AreaConhecimento rdfs:subClassOf owl:Class .
:Matematica rdfs:subClassOf :AreaConhecimento .

# definição do município
:Sao_Paulo a :Município .
:Sao_Paulo dbo:location <http://dbpedia.org/resource/São_Paulo> .
:Sao_Paulo dcterms:title "São Paulo" .
:Sao_Paulo dbo:state <http://dbpedia.org/resource/São_Paulo_(state)> .

# definição dos períodos desejados
:p2014 a :Período .
:p2014 time:years "2014"^^xsd:decimal .
:p2011 a :Período .
:p2011 time:years "2011"^^xsd:decimal .
    
```

Figura 5. Recorte da ontologia para duas das questões de pesquisa (sintaxe Turtle).

Tabela 1. Caracterização dos conjuntos de dados usados para validar a ontologia.

Categoria	Catálogo	Conjunto de dados	Descrição	Resposta à questão de competência (seção 3)
<i>Orçamentário</i>	PDDE - Programa Dinheiro Direto na Escola	Valor repassado pelo programa PDDE aos municípios	Assistência financeira em caráter suplementar, a fim de contribuir para manutenção e melhoria da infraestrutura física e pedagógica.	Total repassado pelo programa PDDE para as escolas do município em 2014: R\$ 57.643.863,23
<i>Censitário</i>	Censo da Educação Básica	Escolas	Cadastro de todas as escolas do país, descrevendo suas características, infraestrutura e de pessoal.	% de escolas municipais com banda larga no Brasil: 70.2%
<i>Acadêmico</i>	SAEB - Sistema Nacional de Avaliação da Educação Básica	Alunos do 9º ano do Ensino Fundamental	Amostra probabilística de escolas públicas e privadas, avaliadas nas disciplinas de Matemática e Língua Portuguesa; contém dados de desempenho e socioeconômico dos alunos (anonimizados); usada para avaliar a evolução do desempenho acadêmico ao longo do tempo	% de escolas amostradas nos níveis socioeconômicos mais altos em 2017: 0,4%
<i>Desempenho</i>	Indicadores sobre	IDEB por município	Compila resultados de dois conceitos igualmente	Nota das escolas municipais de São

	Ensino Básico - IDEB Por Municípios	rede municipal séries finais	importantes para a qualidade da educação: o fluxo escolar e as médias de desempenho nas avaliações. Usado como principal indicador para políticas públicas no ensino básico.	Paulo no IDEB ensino fundamental anos finais em 2011 em Matemática: 4.2
--	-------------------------------------	------------------------------	--	--

6. Discussão e conclusão

A aplicação do modelo de referência demonstrou-se coerente com os macrodados educacionais disponibilizados pelo governo brasileiro, revelando pontos em comum entre as diferentes fontes de dados e permitindo responder questões relevantes acerca destes dados. O estabelecimento de um modelo de referência como este possibilita a modelagem de uma grande rede de dados abertos conectados, uma vez que os dados de diferentes conjuntos estarão anotados semanticamente e conectados entre si por meio da ontologia proposta. Assim, a ontologia populada pode ser usada para o cruzamento de dados dessas diferentes fontes em uma única linguagem, independente dos formatos dos arquivos originais, utilizando padrões da Web como RDF e SPARQL. Penteado et al. (2019) trazem algumas das limitações nos macrodados educacionais atuais. Algumas delas podem ser contornadas com a adoção deste modelo de referência, como os requisitos relacionados ao enriquecimento de dados, vocabulários e a duplicidade de dados. O modelo proposto, embora validado em relação à aplicabilidade, ainda deve ser avaliado em diferentes frameworks de métricas de ontologias como OntoQA (Tartir et al., 2005).

O presente artigo traz como contribuição principal o desenvolvimento de uma ontologia de domínio que unifica as diferentes dimensões dos dados abertos governamentais no domínio macro educacional no Brasil. Embora existam muitos catálogos e conjuntos de dados e suas diferentes distribuições ao longo dos anos, um número reduzido de entidades são definidas e relacionadas entre si para a geração de dados úteis para a sociedade como um todo (pesquisadores, formuladores de políticas públicas, cidadãos e empresas). Em trabalhos futuros, buscaremos refinar esse modelo, para atender a especificidades não contidas nele bem como validá-lo em mais *datasets* ou mesmo a *datasets* de outros países, de modo a mapear sua generalização a outros contextos amplos. De modo secundário, traz também como contribuição um método para extrair ontologias a partir de dados existentes, se beneficiando dos dados abertos já publicados pelos governos nacional e até mesmo em níveis estaduais e federais pelo Brasil – o que, no entanto, ainda necessita de maior validação. Em trabalhos futuros, pode ser testada sua generalização para outros domínios, como saúde, justiça, dados legislativos, dentre os milhares de conjuntos de dados disponíveis atualmente no portal *dados.gov.br*.

Agradecimentos

Os autores agradecem à CAPES - Código de Financiamento 001, CNPq e FAPESP (Processo 2015/24507-2) pelo apoio financeiro dado ao projeto.

Referências

- Berners-Lee, T. (2001). Relational Databases on the Semantic Web. <https://www.w3.org/DesignIssues/RDB-RDF.html>
- Cabral, S., Beduschi, N., Zancanaro, A., Todesco, J., Gauthier, F. (2012). Aplicando Linked Data na Publicação de Dados do ENEM. ONTOBRAS-MOST, vol. 938 of CEURS Workshop Proceedings, p. 176-181, CEUR-WS.org.
- Carchedi, L. C., Barrére, E., Mendonça, F. M., Souza, J. F. (2018). Onto4LA: uma ontologia para integração de dados educacionais. Congresso Brasileiro de Informática na Educação, p. 439-448. <http://dx.doi.org/10.5753/cbie.wcbie.2018.439>.
- Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. Knowledge Acquisition 5: 199-220.
- Gruninger, M. and Fox, M.S. (1995). Methodology for the Design and Evaluation of Ontologies. In: Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI-95, Montreal.
- Guarino, N. (1997). Understanding, Building, and Using Ontologies. International Journal of Human and Computer Studies, v.46, n.2-3, p. 293-310.
- 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, v. 29, n. 4, p. 258–268.
- Noy, N. F., McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Disponível em: https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html
- OASIS. (2014). OASIS SOA Reference Model. Acessado em: 05 jul 2019. Disponível em: <https://www.oasis-open.org/committees/soa-rm/faq.php>.
- Penteadó, 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.
- Penteadó, B. E., Isotani, S. (2017). Dados abertos educacionais: que informações temos disponíveis? VI Congresso Brasileiro de Educação, vol. 4, p. 1933-1938.
- Penteadó, 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.
- 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. 2014.
- Tartir, S., Arpinar, I.B., Moore, M., Sheth, A.P., Aleman-Meza, B. (2005). OntoQA: Metric-Based Ontology Quality Analysis. IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources p. 45-53.