

# Uma proposta de um vocabulário para os dados abertos e conectados das instituições acadêmicas brasileiras

André Moura Lima<sup>1</sup>, Joel da Silva Pereira Filho<sup>1</sup>, Breno Baptista Nahuz<sup>1</sup>,  
Sérgio Souza Costa<sup>1</sup>

<sup>1</sup>Centro de Ciências Exatas e Tecnologias – Universidade Federal do Maranhão  
(UFMA)

Caixa Postal 322 – 65.086-110 – São Luís – MA – Brasil

{am.lima,breno.nahuz,joel.pereira}@discente.ufma.br,sergio.costa@ufma.br

**Abstract.** *In the context of the Access to Information Law (LAI) in Brazil, government transparency has been strengthened, including academic institutions. This article is linked to the DBacademic research project and seeks to create a connected repository of data from federal institutions, requiring a connected data model. 265 collections of data were identified on their own and government portals, highlighting information on courses, teachers and students and others. The use of the Curriculum Course Syllabus Ontology (CCSO) was enhanced to meet national needs, resulting in a detailed ontology, successfully achieving the objective by demonstrating a connected data model.*

**Resumo.** *No contexto da Lei de Acesso à Informação (LAI) no Brasil, a transparência governamental tem sido fortalecida, incluindo instituições acadêmicas. Este artigo, está vinculado ao projeto de pesquisa DBacademic e busca criar um repositório conectado de dados de instituições federais, exigindo um modelo de dados conectado. Foram identificadas 265 coleções de dados em portais próprios e do governo, destacando informações de cursos, docentes e discentes e outros. O uso do Curriculum Course Syllabus Ontology (CCSO) foi aprimorado para atender às necessidades nacionais, resultando em uma ontologia detalhada, alcançando com êxito o objetivo ao demonstrar um modelo de dados conectados.*

## 1. Introdução

Os dados abertos são aqueles que além de públicos, estão em formato adequado, podendo ser reutilizados e redistribuído [PIRES, 2015], já os dados conectados são um conjunto de melhores práticas para publicar e ligar dados na Web [HEATH; BIZER; BERNERS-LEE, 2011] e juntos atendem ambas as propriedades. O acesso aos dados abertos ganhou muito destaque com o movimento de transparência em todo o mundo, podendo trazer benefícios como contribuir com o avanço da ciência a quem publica neste formato [PIRES, 2015]. No Brasil, a Lei nº 12.527, de 18 de novembro de 2011, conhecida como Lei de Acesso à Informação (LAI), juntamente com Instrução Normativa SLTI nº 4, de 13 de abril de 2012, que estabeleceu a Infraestrutura Nacional de Dados Abertos (INDA), bem como o Decreto nº 8.777, de 11 de maio de 2016, que instituiu a Política de Dados Abertos do Poder Executivo Federal, estabelece políticas para publicar dados livres. Como resposta a essa necessidade, as instituições têm disponibilizado uma extensa quantidade de informações públicas e abertas por meio de seus portais. Estabelecendo compromisso do Brasil, no âmbito da Parceria para Governo Aberto (Open Government Partnership - OGP), dentre outros normativos que abordam o tema da transparência pública. Dentre a estas instituições, este trabalho centra-se em analisar as instituições acadêmicas, bem como Universidades Federais e os Institutos Federais de Ciência e Tecnologia (IFETs).

Para atender o decreto, as instituições acadêmicas têm disponibilizado dados abertos por meio de portais. Estes dados têm sido disponibilizados isoladamente e em diferentes formatos e estruturas. Mesmo as instituições acadêmicas tendo diversas coleções em comuns, como discentes, docentes e cursos. Existe uma grande heterogeneidade quanto às estruturas, formatos e mecanismos de acesso aos dados. Essa característica torna impraticável a análise dos dados entre as 108 instituições brasileiras, porque demandaria escrever um *script* específico para cada uma delas. Em [COSTA et. al, 2020], os autores demonstraram como a conexão entre os dados poderia aumentar muito a expressividade das consultas, favorecendo a análise de dados e o desenvolvimento de soluções inovadoras. Essa conectividade é possível via um conjunto de boas práticas propostas por Tim Berners-Lee, denominadas Dados Conectados [BERNES-LEE, 2006]. Em [COSTA et. al, 2021], os autores apresentaram resultados de um protótipo que incluiu 25 instituições acadêmicas, mas destacaram a necessidade de ampliar o estudo para todas as instituições acadêmicas, além da necessidade de elaborar um vocabulário capaz de representar as relações entre os dados destas instituições nacionais.

O objetivo geral deste trabalho é apresentar um modelo de dados conectado para informações acadêmicas de instituições públicas brasileiras. Para isso, será identificado e tabulado os dados abertos disponibilizados pelos portais de todas as universidades e institutos federais. Dentre as diversas coleções de dados, as mais frequentes serão selecionadas para a elaboração do modelo de dados que servirá de base para a elaboração do vocabulário. Espera-se que com o resultado deste trabalho, o projeto DBAcademic [COSTA et. al, 2021], possa ser ampliado para incluir todas as instituições acadêmicas, além de servir para outros trabalhos que necessitem representar os dados de instituições acadêmicas como dados conectados.

## 2. Metodologia

Foram tabulados e coletados dados de 108 instituições brasileiras, incluindo 68 Universidades Federais, 38 Institutos Federais e 2 Centros Federais. A identificação foi feita por meio dos portais próprios das instituições e pelo portal brasileiro de dados abertos mantido pelo governo e documentado em um único repositório. A exemplo, algumas instituições, como a Universidade Federal do Maranhão (UFMA) e o Instituto Federal de Educação, Ciência e Tecnologia de Rondônia (IFRO), disponibilizam dados em seus próprios portais, assim como no portal de dados abertos do governo<sup>1</sup>. Essa pesquisa representou um esforço significativo para catalogar dados de instituições acadêmicas em todo o país.

Dois métodos distintos foram empregados para avaliar a busca de dados. O primeiro se baseou na estrutura de classificação 5 estrelas para dados abertos, concebida por Berners-Lee em 2006. O segundo utilizaram-se alguns princípios que compreendem as diretrizes para dados abertos delineados por Tauberer em 2014. Essas abordagens proporcionaram critérios para avaliar os portais, permitindo a análise do grau de acessibilidade dos dados e seu desenvolvimento, além de revelar as medidas adotadas pelos governos para estruturar as informações divulgadas. O esquema 5 estrelas proposto por Berners-Lee 2006, visa alcançar o nível mais alto da classificação, onde do nível 1 até o nível 5 o esquema, propõem a publicação em formato *Linked Open Data* (LOD) e o uso de vocabulários já existentes. Para Tauberer 2014, os quatorze princípios trata-se de dados abertos governamentais, mas que podem ser associados à busca por dados abertos de instituições acadêmicas. Temos como base o quarto princípio, os dados devem estar disponíveis para uma maior variedade de finalidades.

---

<sup>1</sup> <https://dados.gov.br/>

Na pesquisa foram adotados vocabulários, incluindo o uso de ontologias já existentes como o DC (*Dublin Core*) para metadados genéricos, o FOAF (*Friend Of A Friend*) para relações interpessoais, o schema.org para informações na web, a *Academic Institution Internal Structure Ontology* (AIISO), para representar a estrutura interna de instituições acadêmicas, o (AIISO Roles), para papéis acadêmicos e o CCSO (*Curriculum Course Syllabus Ontology*), para modelar conceitos e entidades em um ambiente educacional, como ementa e currículo. Esses recursos vocabulares facilitam a descrição e relacionamento de dados acadêmicos, contribuindo para a criação de dados conectados e aprimorando a gestão curricular. No entanto, alguns desses vocabulários, como o schema.org, podem apresentar lacunas e incompletudes para descrever os dados de determinadas instituições (Costa et al., 2020; Patel-Schneider, 2014; Styles & Shabir, 2008; Styles & Wallace, 2008; Katis et al., 2018).

### 3. Resultados

O presente artigo irá apresentar os resultados, para a identificação e tabulação dos dados públicos e abertos retirados dos conjuntos de dados, para reutilizar e transformar em dados conectados e possivelmente criação da ontologia para instituições acadêmicas.

#### 3.1. Catalogação

A pesquisa amplia o estudo previamente conduzido por [COSTA et al.,2020], que identificou 25 instituições com os conjuntos de dados mais acessíveis e completos. O presente estudo inclui todas as instituições acadêmicas federais, tais como as universidades e institutos federais de educação tecnológica. Os resultados deste estudo baseiam-se na utilização do censo de educação superior de 2021 para a identificação das instituições de ensino federais. Conforme as informações obtidas através do censo da educação superior, o Brasil possui, atualmente, 68 Universidades Federais, 38 Institutos Federais e 02 Centros Federais de Educação Tecnológica (Cefet), além da Universidade Tecnológica Federal do Paraná (UTFPR), somando um total de 108 entidades acadêmicas que estão sob a responsabilidade do governo federal e que tiveram de se ajustar à Política de Dados Abertos do Poder Executivo Federal. A Tabela 1 apresenta como os dados estão sendo publicados pelas universidades federais.

**Tabela 1. Abordagem de publicação dos dados nas Universidades**

<b>Publicação dos dados</b>	<b>Quantidade</b>
Não foram encontrados dados abertos	16
Possuíam portal próprio para disponibilização dos dados abertos	43
Possuíam portal próprio para disponibilização dos dados abertos, porém estava fora do ar nos dias consultados	3
Disponibilizavam seus dados somente pelo portal de dados abertos do governo federal	7
Disponibilizavam seus dados pelo portal de dados abertos do governo federal e seu próprio portal	27
Disponibilizam separadamente grupos de dados abertos em seus portais	36
Disponibilizam separadamente grupos de dados abertos em seus portais, porém não possui nada publicado	1

Na análise realizada, a maioria das universidades possui portais próprios para os dados abertos, embora alguns tiveram problemas de acessibilidade não abordados até então. Além de portais próprios, algumas instituições compartilham seus dados com o portal do governo federal, sendo apenas sete provedores exclusivos deste. Ademais, 16 universidades não disponibilizaram seus dados publicamente, incluindo a Universidade Federal do Acre e a Universidade Federal de Alagoas, embora algumas já tenham iniciado planos de dados abertos. Apenas um número limitado de organizações possui grupos dedicados à publicação de dados, sendo o exemplo mais notável aqueles relacionados ao ensino, pesquisa e extensão.

O desafio principal reside na quantidade de dados oferecidos. Entre as universidades, algumas disponibilizam apenas um conjunto de dados. A Tabela 2, oferece uma síntese dessa variedade de disponibilidade de dados.

**Tabela 2. Conjuntos de dados disponíveis nas Universidades**

<b>Conjunto de dados disponível</b>	<b>Resultado</b>
Quantidade média de conjuntos de dado publicados	24
Maior quantidade de conjunto de dados publicados	224
Menor quantidade de conjunto de dados publicados	1

Apesar do Brasil possuir 68 Universidades Federais, há uma pouca disponibilidade de dados. Atualmente, das instituições analisadas, mais de 30 têm conjuntos de dados, mas, apenas a Universidade Federal do Mato Grosso, encontra-se com maior disponibilidade, com 224 conjuntos de dados publicados. Com relação aos institutos federais de ensino, a Tabela 3 apresenta como estas instituições estão disponibilizando os dados abertos.

**Tabela 3. Abordagem de publicação dos dados nos Institutos**

<b>Publicação dos dados</b>	<b>Quantidade</b>
Não foram encontrados dados abertos	1
Possuíam portal próprio para disponibilização dos dados abertos	30
Possuíam portal próprio para disponibilização dos dados abertos, porém, estava fora do ar nos dias consultados	2
Disponibilizavam seus dados pelo portal de dados abertos do governo federal	11
Disponibilizavam seus dados pelo portal de dados abertos do governo federal e seu próprio portal	8
Disponibilizam separadamente grupos de dados abertos em seus portais	18

Similar às universidades, muitos institutos enfrentam desafios na disponibilização de dados abertos, incluindo a publicação de apenas um conjunto limitado de informações, como evidenciado na Tabela 4. Esses desafios são indicativos

de questões comuns relacionadas à transparência de dados que afetam tanto as instituições de ensino superior quanto os institutos, ressaltando a necessidade de melhorias nessa área em ambas as esferas acadêmicas.

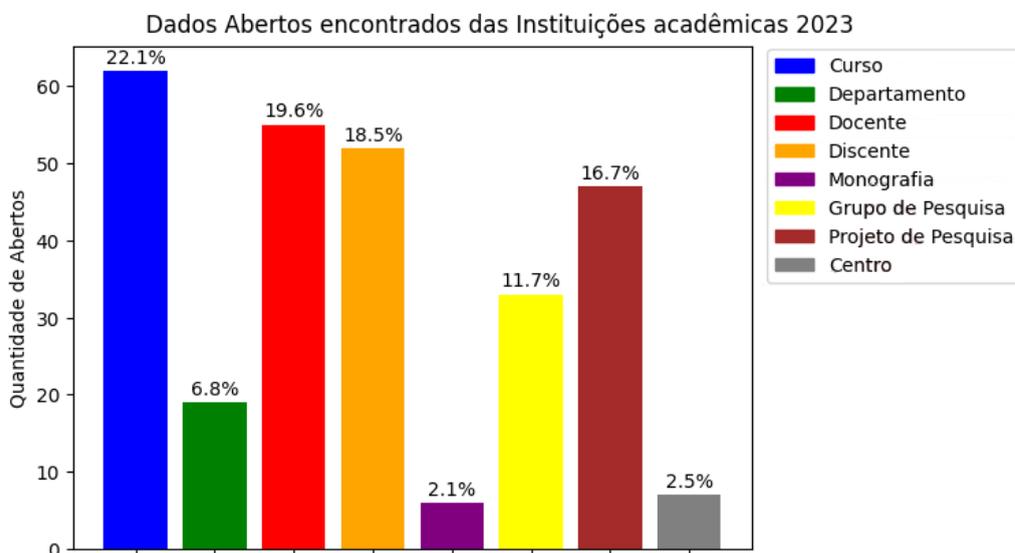
**Tabela 4. Conjuntos de dados disponíveis nos Institutos**

Conjunto de dados disponível	Resultado
Quantidade média de conjuntos de dados publicados	20
Maior quantidade de conjunto de dados publicados	93
Menor quantidade de conjunto de dados publicados	1

A maior média de publicação de dados é do Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA) contando que cada instituto tem um dado publicado com mais de 30 conjuntos de dados enviados. Tanto esses dados dos institutos e universidades ainda podem ser revisados e estudados mais adiante.

### 3.2. Seleção dos conjuntos de dados

A seleção dos dados abertos que serão utilizados nas próximas etapas foi feita através da identificação, categorização e tabulação dessas informações extraídas por categoria nos portais dessas instituições acadêmicas. Após a análise dos portais de dados abertos, identificaram-se as seguintes categorias mais comuns: cursos, departamentos, docentes, discentes, monografias, grupos de pesquisa, projetos de pesquisa e centros. Essas categorias oferecem detalhes abrangentes sobre as atividades acadêmicas e de pesquisa das instituições. A tabulação indicou também quais sites estavam disponíveis, tornando-as adequadas para a criação de dados conectados. Em 2023, o percentual de dados abertos extraídos foi determinado com base na quantidade de dados em cada categoria conforme mostrado na Figura 1.



**Figura 1. Dados abertos de instituições acadêmicas.**

Os dados extraídos revelam que várias instituições acadêmicas no Brasil, disponibilizam informações em formato de dados abertos, mas a disponibilidade varia por categoria, com destaque para Cursos, Docentes e Discentes. Esses dados podem ser acessados por meio do Portal Brasileiro de Dados Abertos, de uma ferramenta de

visualização da CGU, Painel de monitoramento de dados abertos e no portal próprio da instituição.

Os formatos mais comuns para acessar esses são JSON, CSV, ODS e HTML/PDF para dicionários de dados. No entanto, alguns dados não estavam disponíveis ou estavam desatualizados. A análise ressalta a importância da publicação de dados abertos pelas instituições acadêmicas para atender a requisitos legais e permitir sua reutilização. Em várias universidades e institutos, a pesquisa de dados relacionados a professores, alunos e outros conjuntos de informações revelou que alguns desses dados não puderam ser localizados. Os conjuntos de dados mais comuns encontrados são detalhados na Tabela 5.

**Tabela 5. Quantidade por conjunto de dados e tipo instituição acadêmica**

Conjunto de dados	Universidades	Institutos	Total
Curso	27	30	57
Docente	30	22	52
Discente	31	19	50
Departamento	9	10	19
Centro	5	2	7
Grupo de pesquisa	15	10	25
Projeto de pesquisa	24	25	49
Monografia	3	3	6
Total			265

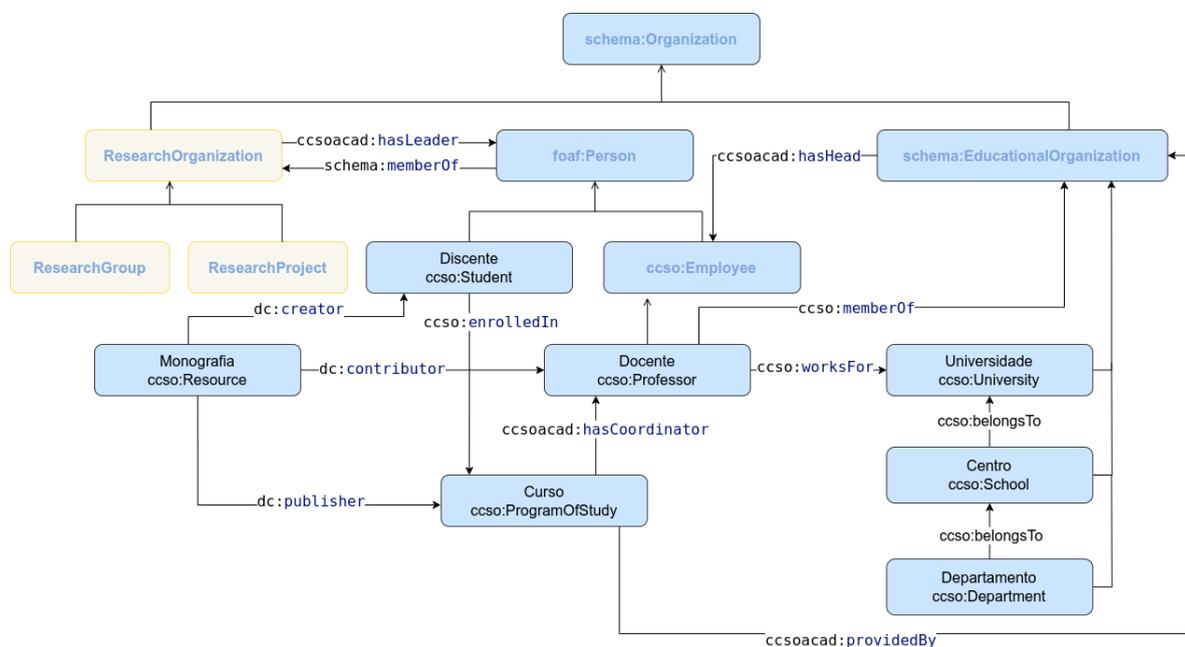
Na análise de agosto de 2023, foram identificados 265 conjuntos de dados de várias instituições. Cada conjunto foi avaliado quanto à data de criação, última atualização, URLs da página e dos dados, formato dos dados e atributos disponíveis. Esses conjuntos de dados são os dados disponibilizados por cada instituição, que incluem informações sobre servidores, estudantes, projetos de pesquisa, despesas e orçamentos e outros. Eles são indexados pelo portal brasileiro de Dados Abertos e pelo próprio portal da instituição, cada conjunto de dados tem suas próprias informações, por exemplo, o curso tem informações como nome, campus, tipo de curso, departamento, unidade, informações essas que foram importantes para saber a quantidade de dados abertos que as instituições acadêmicas disponibilizam que foi fundamental para criação da ontologia. Esses dados são essenciais para acompanhar a disponibilidade e a relevância das informações fornecidas pelas instituições, promovendo a transparência e o uso eficaz dos dados abertos.

### 4.3. Elaboração e reuso de vocabulários

Como destacado na metodologia, para alcançar as cinco estrelas os dados devem se conectar, sempre que possível, a outros dados e vocabulários já existentes [BERNERS-LEE, 2006]. Dessa maneira, um dos resultados esperados neste trabalho é a identificação dos vocabulários já existentes que seria suficiente para representar os dados extraídos das coleções de dados. Com reuso de vocabulário, o resultado de dados conectados se deu a partir do relacionamento de classes e propriedades dos vocabulários como FOAF (*Friend of a friend*), DC (*Dublin Core*) e *Schema.org*. Além disso, o CCSO

foi comparado com os vocabulários (AIISO) *Academic Institution Internal Structure Ontology* e *Academic Institution Internal Structure Ontology Roles (AIISO Roles)*, para a transformação desses dados. Assim, (CCSO) *Curriculum Course Syllabus Ontology*, associado também ao *Schema.org*, possibilitou a criação da ontologia e quanto à elaboração do vocabulário para a criação da ontologia, podemos citar a criação de dados conectados para instituições acadêmicas. Em [COSTA et al., 2020], foi apresentado um modelo para criação e transformação desses dados que está sendo ampliada e discutida neste projeto.

A ontologia elaborada teve como base o reuso dos dados abertos extraídos de cada categoria, conforme mostrado na subtópico coleções de dados, pesquisa essa realizada no decorrer do projeto, que tem como base modelo de dados conectados para informação acadêmica de instituições públicas brasileiras. Além disso, a ontologia desenvolvida não apenas demonstra uma abordagem inovadora para a representação de informações acadêmicas de instituições públicas brasileiras, mas também oferece um potencial significativo para aprimoramentos e expansões futuras. A flexibilidade inerente ao modelo de dados conectados permite a incorporação de novos atributos, relacionamentos e fontes de dados, garantindo que a ontologia possa se adaptar às evoluções do ambiente acadêmico e às demandas em constante mudança.



**Figura 2. Uma extensão do vocabulário CCSO**

O esquema de ontologia apresentado na Figura 2, descreve uma estrutura taxonômica para organizações. No entanto, o vocabulário utilizado foi o CCSO (*Curriculum Course Syllabus Ontology*) para definir a relação entre cada classe e propriedade da organização, a definição completa pode ser acessada no endereço eletrônico<sup>2</sup>. Essa ontologia define os seguintes relacionamentos para dados conectados conforme a Tabela 6.

<sup>2</sup> <https://lambdageo.github.io/ccso-dbacademic/>

Tabela 6. Relacionamento para vocabulário CCSO

<b>Relacionamento da Ontologia</b>		
<b>Classes:</b> Research Group <sup>C</sup> , Research Group <sup>C</sup> , Research Project <sup>C</sup> .		
<b>Propriedades:</b>		
<b>hasLeader:</b> (Tem líder), faz o relacionamento entre uma entidade e uma Organização.	<b>hasCoordinator:</b> (Tem Coordenador), estabelece uma relação entre uma entidade e outra entidade que representa o coordenador.	<b>providedBy:</b> (Fornecido por), representa um programa de estudo fornecido por uma instituição.
<b>memberOf:</b> membro de algo.	<b>enrolledIn:</b> representa discente matriculado no curso.	<b>worksFor:</b> representa docente que trabalha no departamento.
<b>dc:creator:</b> representa o criador da monografia.	<b>dc:contributor:</b> representa o contribuidor da monografia.	<b>dc:publisher:</b> representa uma entidade responsável por disponibilizar o recurso como uma monografia.
<b>belongsTo:</b> representa a departamento e centro pertence à instituição.		

Com o uso do CCSO esse relacionamento é feito entre as categorias analisadas com suas classes e propriedade, docente, discente, departamento, centro e universidade, fazendo o ligamento entre os dados. Podemos ampliar os resultados, para definir cada atributo das categorias em dados conectados. Dessa forma, a estrutura complementar da ontologia proposta neste projeto e iniciada por Costa et al.(2020) pode ser encontrada no repositório GitHub<sup>3</sup>. Onde está relacionada com a coleção de dados, fazendo assim uma proposta para os dados conectados das instituições públicas brasileiras elaborando mais triplas. Ao expandir a quantidade de triplas, é possível criar relações mais complexas entre os dados, aumentando a capacidade de inferência e análise semântica. Vale também lembrar que a integração e a disponibilização desses dados provenientes de instituições públicas brasileiras têm ganhado crescente importância na era da informação. A proposta de estabelecer uma coleção de dados conectados visa não apenas centralizar informações, mas também promover a interoperabilidade e a reutilização desses dados para diversos fins. Essa iniciativa pode ser efetivada por meio da adoção de padrões abertos e estruturas semânticas, como o uso da tecnologia *Linked Data* e da modelagem em RDF (*Resource Description Framework*), recomendadas pelo *World Wide Web Consortium* (W3C).

## 5. Considerações Finais

Os resultados deste trabalho incluiu uma análise de todos os dados abertos das instituições acadêmicas brasileiras além do vocabulário desenvolvido como uma extensão ao vocabulário CCSO. A pesquisa abrangeu instituições sob a Lei de Acesso à Informação (LAI), a Instrução Normativa SLTI nº 4/2012 e o Decreto 8.777/2016, que

<sup>3</sup> <https://lambdageo.github.io/projects/dbacademic/modelagem/>

exigiam a criação de Planos de Dados Abertos (PDA). Observou-se que todas as instituições possuem PDAs, mas muitas enfrentam problemas de disponibilidade e atualização de dados em seus portais.

O objetivo principal do trabalho foi alcançado e a extensão proposta foi publicada e utilizada em um trabalho recentemente publicado. O vocabulário estendido foi capaz de representar as principais propriedades e relação entre as 8 principais coleções de dados utilizadas neste trabalho. Esse vocabulário está disponível no GitHub e poderá ser utilizado e revisado em trabalhos futuros. Além do objetivo principal, este trabalho foi capaz de demonstrar que mesmo atendendo o Decreto 8.777/2016, ainda é necessário um avanço maior na direção de políticas para o uso e avaliação dos dados abertos atualmente disponíveis por estas instituições.

## Referências

- ALENCAR, André de et al. Publicação e consumo de dados abertos conectados acadêmicos. *Revista Principia - Divulgação Científica e Tecnológica do IFPB, João Pessoa*, n. 42, p. 136-145, ago. 2018. ISSN 2447-9187. Disponível em: <https://periodicos.ifpb.edu.br/index.php/principia/article/view/1988>. Acesso em: 20 Ago. 2023. doi:<http://dx.doi.org/10.18265/1517-03062015v1n42p136-145>.
- BERNERS-LEE, T. *Linked Data - Design Issues*: 2006. Disponível em:<https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 14 jul. 2023.
- BIZER, C; HEATH, T; BERNERS-LEE, T. *Linked data: The story so far*. In: *Semantic services, interoperability and web applications: emerging concepts*. IGI global, 2011. p. 205-227. Disponível em: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>. Acesso em: 21 mar. 2023.
- BRASIL, Ministério do Planejamento, Orçamento e Gestão. *Instrução Normativa nº 4, de 19 de abril de 2012*. 2012. Disponível em: [https://www.gov.br/agricultura/pt-br/acesso-a-informacao/dadosabertos/arquivos-raiz/in04\\_2012.pdf](https://www.gov.br/agricultura/pt-br/acesso-a-informacao/dadosabertos/arquivos-raiz/in04_2012.pdf). Acesso em: mar.2023.
- BRASIL. Decreto no 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal. 11 maio de 2016. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2016/decreto/d8777.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm). Acesso em: 17 maio. 2023.
- BRASIL. Lei no 12.527, de 18 de novembro de 2011 [Lei de Acesso à Informação]. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. 2011. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm). Acesso em: 17 maio. 2023.
- BRASIL. Portal Brasileiro de Dados Abertos. 2023. Disponível em: <https://dados.gov.br>. Acesso em: 23 jan. 2023.
- COSTA, S. S.; SOUSA, M. V. D.; SILVA, M. L. da; OLIVEIRA, E. C. de; GUIMARÃES, J. V. M. Uma solução semi-automática para extração, transformação e carga de dados abertos conectados. In: *WORKSHOP DE INFORMAÇÃO, DADOS E TECNOLOGIA, 2019*. Anais [...]. Brasília: FCI, 2019. p.138-143. Disponível em: <http://widat2019.fci.unb.br/index.php/anais-widat-2019>. Acesso em:

21 mar. 2023.

- COSTA, S. S et al. DBacademic: Conectando os dados abertos das instituições de ensino do Brasil. *Ciência da Informação*, v. 49, n. 3, 2020.
- INEP. Censo de educação superior 2021 - Divulgação de resultados. Ministério da Educação, 2022. Disponível em: [https://download.inep.gov.br/educacao\\_superior/centso\\_superior/documentos/2021/apresentacao\\_censo\\_da\\_educacao\\_superior\\_2021.pdf](https://download.inep.gov.br/educacao_superior/centso_superior/documentos/2021/apresentacao_censo_da_educacao_superior_2021.pdf). Acesso em: 14 mar. 2023.
- KATIS, E.; KONDYLAKIS, H.; AGATHANGELOS, G.; VASSILAKIS, K. Developing an ontology for curriculum and syllabus. In: *Extended Semantic Web Conference*. [S.l.: s.n.], 2018.
- KESSLER, C.; KAUPPINEN, T. Linked open data university of münster—infrastructure and applications. In: SPRINGER. *Extended Semantic Web Conference*. 2012. p. 447–451. Disponível em: [https://link.springer.com/content/pdf/10.1007/978-3-662-46641-4\\_43.pdf?pdf=inline%20link](https://link.springer.com/content/pdf/10.1007/978-3-662-46641-4_43.pdf?pdf=inline%20link). Acesso em: 26.maio.2023.
- NAHUZ, Breno Baptista et al. Uma Abordagem Baseada Em Engenharia De Dados Para Extração, Transformação E Carregamento De Dados De Instituições Acadêmicas. In: *III Simpósio REACT sobre Descarbonização: economia, energia e ambiente*. Anais. São Luís(MA) Online, 2023.
- PATEL-SCHNEIDER, Peter F. Analyzing schema. Org. In: *The Semantic Web—ISWC 2014: 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I 13*. Springer International Publishing, 2014. p. 261-276.
- PIRES, M.T. Guia de dados abertos. São Paulo: Governo do Estado de São Paulo, 2015. Disponível em: [https://nic.br/media/docs/publicacoes/13/Guia\\_Dados\\_Abertos.pdf](https://nic.br/media/docs/publicacoes/13/Guia_Dados_Abertos.pdf). Acesso em: 23.jun.2023.
- ROCHA, J; LÓSCIO, B. OpenCIn: Usando Dados Abertos e Conectados para a Publicação de dados sobre o CIn/UFPE. In: *CONCURSO DE TRABALHOS DE INICIAÇÃO CIENTÍFICA DA SBC (CTIC-SBC)*, 34. , 2015, Recife. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2015. p. 11-20. Disponível em: <https://sol.sbc.org.br/index.php/ctic/article/view/10014/9898>. Acesso em: 20 Ago.2023.
- STYLES, R.; SHABIR, N. Academic institution internal structure ontology (aiiso). 2008. Disponível em: <https://vocab.org/aiiso/>. Acesso em: 19 jun.2023.
- STYLES, R.; WALLACE, C. Academic institution internal structure ontology Roles (AIISO Roles). 2008. Disponível em: <https://vocab.org/aiiso-roles/>. Acesso em: 19 jun.2023.
- TAUBERER, J. 14 Principles of Open Government Data - Open Government Data: TheBook. 2nd ed.2014. Disponível em: <https://opengovdata.io/2014/principles/>. Acesso em: 21 mar. 2023.