

Dados abertos dos Institutos Federais de Educação: análise de temas sugeridos pela Controladoria-Geral da União

Open data from the Federal Institutes of Education: analysis of themes suggested by the Office of the Comptroller General

Yuri Bento Marques¹, Marcello Peixoto Bax², Geraldo Lopes Júnior³, Gustavo Henrique Silva de Souza⁴, Eduardo Cardoso Melo⁵

¹ Instituto Federal do Norte de Minas Gerais (IFNMG), Belo Horizonte, MG, Brasil. ORCID: <https://orcid.org/0000-0003-3414-2325>

² Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil. ORCID: <https://orcid.org/0000-0003-0503-3031>

³ Instituto Federal do Norte de Minas Gerais (IFNMG), Belo Horizonte, MG, Brasil. ORCID: <https://orcid.org/0000-0002-2243-2336>

⁴ Instituto Federal do Norte de Minas Gerais (IFNMG), Belo Horizonte, MG, Brasil. ORCID: <https://orcid.org/0000-0002-4046-9669>

⁵ Instituto Federal de Minas Gerais (IFMG), Belo Horizonte, MG, Brasil. ORCID: <https://orcid.org/0000-0002-1323-5859>

Autor para correspondência/Mail to: Yuri Bento Marques, yuri.marques@ifnmg.edu.br

Recebido/Submitted: 25 de agosto de 2021; **Aceito/Approved:** 12 de outubro de 2021



Copyright © 2022 Marques, Bax, Lopes Júnior, Souza & Melo. Todo o conteúdo da Revista (incluindo-se instruções, política editorial e modelos) está sob uma licença Creative Commons Atribuição 4.0 Internacional. Ao serem publicados por esta Revista, os artigos são de livre uso em ambientes educacionais, de pesquisa e não comerciais, com atribuição de autoria obrigatória. Mais informações em <http://revistas.ufpr.br/atoz/about/submissions#copyrightNotice>.

Resumo

Introdução: A abertura das bases de dados governamentais pode contribuir positivamente com a democracia, transparência e controle social, pois oportuniza ao cidadão o acesso, reutilização, modificação e redistribuição livre das informações não sigilosas dos órgãos públicos. Porém, nem todos os dados de interesse da sociedade se encontram publicados em formatos abertos no presente momento. Esse estudo verifica se os dados referentes aos Institutos Federais de Educação, Ciência e Tecnologia, que são comumente solicitados à Controladoria-Geral da União pela sociedade civil, estão sendo catalogados no Portal Brasileiro de Dados Abertos. **Método:** Foi conduzida uma pesquisa quantitativa no Portal Brasileiro de Dados Abertos com suporte da técnica de pesquisa documental, o que permitiu analisar os conjuntos de dados publicados pelos Institutos Federais. Criou-se um índice para analisar cada um dos 14 temas coletados, o qual originou um indicador quantitativo utilizado na comparação dos resultados de cada Instituto Federal envolvido na pesquisa. **Resultados:** Identificou-se que 73,70% dos Institutos Federais catalogaram pelo menos um conjunto de dados no Portal Brasileiro de Dados Abertos. No entanto, verificaram-se contundentemente diversas lacunas na disponibilização de dados abertos por parte dessas instituições, além da baixa aderência à abertura de dados em conformidade às recomendações da Controladoria-Geral da União. **Conclusão:** Estudos como o realizado neste artigo contribuem para que a Administração Pública, especialmente considerando o contexto dos Institutos Federais, tenham subsídios que indiquem possíveis pontos de melhorias no que tange às demandas da sociedade sobre transparência e governança.

Palavras-chave: Acesso à Informação Pública. Dados Governamentais Abertos. Transparência. Institutos Federais de Educação.

Abstract

Introduction: Opening government databases can positively contribute to democracy, transparency and social control, as it provides opportunities for citizens to access, reuse, modify and redistribute non-confidential information from public bodies. However, not all data of interest to society are currently published in open formats. This study verifies whether the data referring to the Federal Institutes of Education, Science and Technology, which are commonly requested from the Office of the Comptroller General by civil society, are being cataloged in the Brazilian Open Data Portal. **Method:** Quantitative research was conducted on the Brazilian Open Data Portal with the support of the documentary research technique, which allowed us to analyze the datasets published by the Federal Institutes. An index was created to analyze each of the 14 collected themes, which originated a quantitative indicator used to compare the results of each Federal Institute involved in the research. **Results:** It was identified that 73.70% of the Federal Institutes cataloged at least one dataset on the Brazilian Open Data Portal. However, there were strikingly several gaps in the availability of open data by these institutions, in addition to the low adherence to opening data in accordance with the recommendations of the Office of the Comptroller General. **Conclusions:** Studies such as the one carried out in this paper contribute to the Public Administration, especially considering the context of Federal Institutes, to have subsidies that indicate points for improvement with regard to society's demands on transparency and governance.

Keywords: Access to Public Information. Open Government Data. Transparency. Federal Institutes of Education.

INTRODUÇÃO

A Assembleia Geral das Nações Unidas reconheceu em 1946 o acesso à informação como um direito humano universal, assim, todas as pessoas têm o direito fundamental de coletar, transmitir e publicar informações. Visando a possibilitar a transparência e condução aberta de suas ações, os governos devem viabilizar o acesso às suas informações. Nos anos 1980, em meio ao movimento do Software Livre, surgiu a Cultura Aberta ou Movimento Aberto a fim de dar suporte à Transparência e Acesso à Informação (Cardoso, 2019, p. 32).

Tanto na esfera pública quanto privada, não é novidade que as organizações precisam gerenciar seus recursos de informação para melhorar sua produtividade, competitividade e desempenho. Para favorecer a adoção e implementação de recursos que auxiliem tal gerenciamento, as instituições precisam estudar e reconhecer quais

fatores organizacionais e culturais contribuem para este processo Bergeron (1996). Uma organização que consiga integrar a criação de conhecimento, sentidos e a tomada de decisões está em vantagem, pois consegue realizar ações baseadas em uma compreensão compartilhada do contexto, com inteligência, criatividade e até mesmo astúcia Choo (1996).

Especificamente na esfera pública, a aplicação dos conceitos de dados governamentais abertos (ou OGD, *Open Government Data*) faz com que as organizações disponibilizem suas bases de dados na *Web* em formato aberto e estruturado. Essa abertura contribui positivamente para o fortalecimento da democracia, transparência e controle social, pois oportuniza ao cidadão o acesso, reutilização, modificação e redistribuição livre das informações não sigilosas dos órgãos públicos Possamai (2016). Adicionalmente, os OGDs colaboram para que as instituições públicas maximizem a integração e a interoperabilidade dos seus dados, obtenham novas informações para subsidiar a tomada de decisão, assim como promover colaboração governo-sociedade na criação de novos serviços e aplicações a partir dos dados abertos Possamai (2016). Para dar suporte à interoperabilidade e ao reuso dos dados, o W3C (W3C, 2017) recomenda reutilizar Ontologias/Vocabulários consolidados, objetivando aumento do consenso sobre os conceitos e termos entre os usuários e a diminuição da redundância e da ambiguidade.

No entanto, a ausência de padrões para a publicação de dados abertos governamentais prejudica a transparência, a análise e a compreensão por parte dos cidadãos. Ademais, poucos dados de interesse da sociedade se encontram publicados em formatos abertos e de fácil acesso. Como consequência disso, são gerados novos pedidos de acesso à informação via Lei de Acesso à Informação Gonçalves (2017), eventualmente sobrecarregando os envolvidos, na medida em que precisam reorganizar sua rotina de trabalho para atender à demanda de informação recebida. Há de se considerar, ainda, um importante aspecto tratado por Monteiro e Lucas (2019) referente ao prazo de vida dos dados disponibilizados, pois a oferta em um ambiente compartilhado precisa ser acompanhada de um planejamento para que ocorra sua constante atualização de forma a atender continuamente as demandas da sociedade.

Ausência de estruturação, baixa semântica e indisponibilidade de informações de interesse da sociedade em OGD é comum aos Institutos Federais de Educação, Ciência e Tecnologia. Conforme evidenciado por Klein (2018, p. 77), nenhum Instituto Federal observa todos os critérios de transparência definidos pela legislação de boas práticas em seus planos de dados abertos. De acordo com Nunes (2018), a Rede Federal de Educação Profissional, Científica e Tecnológica (RFEPCT) é composta por trinta e oito Institutos Federais de Educação, Ciência e Tecnologia, criados pela Lei nº 11.892/2008: Universidade Tecnológica Federal do Paraná, Centros Federais de Educação Tecnológica Celso Suckow da Fonseca e de Minas Gerais, vinte e duas Escolas Técnicas Vinculadas às Universidades Federais e pelo Colégio Pedro II. Além disso, o fato de os Institutos Federais serem compostos por uma estrutura *multicampi*, permite que a área de abrangência de ações, oferta de cursos e fortalecimento dos arranjos produtivos, sociais e culturais seja ampliada.

Tem-se evidenciado um crescimento dos estudos focados na sistematização de informações por meio de dados semânticos, como os estudos de Britto, Ruy, e Azevedo (2020) sobre integração semântica de despesas públicas, Chen et al. (2020) sobre representação ontológica das proteínas e Zaidan e Bax (2013) quanto à agregação de valor aos dados aplicando a tecnologia semântica de *Linked Open Data* (LOD). A relevância disso ancora a necessidade latente de investigar quais as contribuições das tecnologias semânticas aplicadas no contexto de OGD dos Institutos Federais.

Aleixo (2020) afirma que, embora as pesquisas acerca de dados abertos governamentais estejam em evidência atualmente, não estão sendo analisadas como as tecnologias impactam no acesso e disponibilização dos dados. A autora acrescenta ainda que a própria legislação brasileira não é clara acerca dos procedimentos que devem ser realizados para garantir o acesso e publicação dos dados abertos. Isto posto, a complexa estrutura da RFEPCT com suas 661 unidades, demanda uma gestão aberta e pautada em transparência e responsabilidade. O objetivo deste trabalho é analisar as bases de dados abertas por Institutos Federais em relação aos temas sugeridos pela Controladoria-Geral da União (CGU), tais como a publicação de informações sobre alunos, cursos, disciplinas, dentre outros, com o intuito de identificar se esses dados estão devidamente catalogados no Portal Brasileiro de Dados Abertos. Espera-se auxiliar a Administração Pública a promover as adequações necessárias e criar um ambiente acessível e com maior transparência.

O restante do artigo é organizado como segue: na Seção 2, são apresentados os conceitos teóricos envolvendo dados abertos governamentais que suportam este artigo, bem como estudos realizados por outros pesquisadores da área. A metodologia utilizada para a construção deste trabalho é explicada na Seção 3, assim como as técnicas empregadas para a obtenção dos resultados. Na Seção 4 são disponibilizados e discutidos os resultados obtidos pelo artigo. Por fim, na Seção 5 são feitas as considerações finais do estudo.

FUNDAMENTAÇÃO TEÓRICA

Web Semântica

A Web Semântica é considerada uma extensão da *World Wide Web* moderna, na qual os documentos são explicados para as máquinas através de metadados e técnicas normalizadas semanticamente. Desta forma, tanto os usuários humanos quanto as máquinas usufruem de dados mais abertos, estruturados e conectados [Alencar, Xavier, Chaves, e Souza \(2018\)](#).

A linguagem *Hypertext Markup Language* (HTML) foi desenvolvida para humanos consumirem o conteúdo dos sites da Internet, mas não atende a esse objetivo para as máquinas. Para um processamento mais sofisticado dos sites por máquinas, utilizam-se vários modelos para estruturar os dados. O modelo formal de dados e sintaxe RDF (*Resource Description Framework*) está entre os mais populares. Ele viabiliza a criação de metadados que habilitam aplicações da Web Semântica a consumirem de forma autônoma, as informações dos sites [Breitman \(2005\)](#).

A falta de um processo sistemático para criar e reutilizar ontologias que descrevam as informações contidas em dados abertos é um dos principais desafios à interoperabilidade dos sistemas [Santarem Segundo \(2015\)](#).

Dados Abertos Governamentais

De acordo com [Silva \(2018\)](#), apesar da Constituição da República, nos artigos 5º, 37 e 216, já introduzir o acesso à informação pública, a política de informação como política de governo aos documentos governamentais ficou em evidência apenas a partir de 2011, devido à adesão do Brasil ao *Open Government Partnership* (OGP) e à aprovação da Lei nº 12.527 de 2011, mais conhecida como Lei de Acesso à Informação (LAI). A OGP tem como objetivo fomentar, em nível global, práticas de transparência orçamentária dos governos, assim como o acesso às informações e participação da sociedade. A LAI tem como propósito garantir o direito fundamental de acesso à informação acerca dos dados da administração pública.

A fim de assegurar e facilitar o acesso da sociedade aos dados gerados ou custodiados pelo Poder Executivo federal, foi instituída a Infraestrutura Nacional de Dados Abertos (INDA) por meio da Instrução Normativa nº 4 de 12 de abril de 2014. Compete a ela divulgar materiais com orientações para a abertura de dados e elaborar o Plano de Dados Abertos (PDA) das informações públicas [Silva \(2018\)](#).

Com exceção dos dados sigilosos, os órgãos públicos devem disponibilizar em formato aberto todas as suas bases de dados. Porém, essa abertura necessita de estruturação, hospedagem e integração dos dados, o que talvez origine um processo complexo e custoso. Diante disso, é necessário o envolvimento da sociedade e a aplicação de uma matriz de prioridade para a definição de quais bases de dados precisam ser disponibilizadas. Ademais, para que as ações cotidianas realizadas pela entidade governamental visem à abertura das suas respectivas bases de dados, é preciso ser cultivada no órgão a cultura de dados abertos, bem como a realização de atividades voltadas ao engajamento dos envolvidos. Por fim, o apoio contínuo do setor de Tecnologia da Informação no processo de elaboração do PDA auxilia a incorporação da abertura das bases de dados, até mesmo durante a etapa de desenvolvimento dos softwares da instituição ([CGU, 2020](#)).

Foram encontrados diversos trabalhos na literatura que investigam a disponibilização de dados abertos governamentais. Diante disso, apenas aqueles estudos que mais se aproximam da investigação proposta neste artigo são apresentados a seguir.

[Klein \(2018\)](#) investigou os níveis de transparência ativa e passiva, boas práticas e plano de dados abertos de Universidades e Institutos Federais. Para atingir esse objetivo, a pesquisadora criou um indicador de transparência baseado na legislação brasileira e indicadores de transparência da Controladoria Geral da União (CGU) e Ministério Público Federal (MPF). Destacam-se as seguintes conclusões da pesquisadora após a aplicação do indicador: (1) nenhuma Universidade Federal ou Instituto Federal é 100% transparente; (2) quanto maior o orçamento por matrícula, a transparência tende a diminuir; (3) quanto maior o nível de transparência, menores são as reclamações e pedidos de informações e recursos.

[Carossi, Pedan, e Teixeira Filho \(2020\)](#) investigaram os temas mais indagados no Sistema Eletrônico do Serviço de Informações ao Cidadão para as Instituições Federais de Ensino Superior públicas do Brasil. Essa pesquisa utilizou como metodologia a *Grounded Theory* (Teoria Fundamentada em Dados) e resultou na identificação das categorias e temas que seriam prioritários para subsidiar os Planos de Dados Abertos destas instituições.

[Lima, Cunha, e Maior Filho \(2017\)](#) realizaram um estudo de caso no Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB), a fim de identificar as fragilidades e potencialidades do Governo Aberto na instituição. Como resultados, descobriram que embora o IFPB disponibilize para a sociedade diversos dados orçamentários, por outro lado, alguns desses dados são difíceis de localizar no site e, devido a formatos inadequados, difíceis de reutilizar.

Lemos (2017) identificou a falta de maturidade e padronização no processo de abertura e manutenção de dados abertos em Institutos Federais. Diante disso, propôs um processo para publicação de OGD para a RFEPCT fundamentado em Gerenciamento de Processos de Negócios (ou BPM, *Business Process Management*), com o emprego da notação BPMN (*Business Process Model and Notation*).

Silva (2018) desenvolveu uma métrica do reuso dos dados governamentais abertos disponibilizados por Instituições Públicas Federais brasileiras. Então, como resultado da aplicação da métrica em um órgão da administração pública federal, a autora concluiu que o reuso dos dados abertos era incipiente em relação à quantidade de dados disponibilizados.

Quanto à disponibilização dos dados abertos por Institutos Federais de Educação, Ciência e Tecnologia, de acordo com Nunes (2018), a maior motivação para que essas instituições elaborassem um Plano de Dados Abertos (PDA) foi evitar que órgãos de controle realizassem a cobrança acerca da disponibilização de tais dados. Ademais, a ausência de sistemas de Tecnologia da Informação e sensibilização dos servidores públicos da instituição são fatores que dificultam a elaboração e implantação do PDA.

Macedo e Lemos (2021) afirmam que a literatura carece de estudos científicos, sejam eles teóricos ou práticos, que ampliem o entendimento sobre qual é o atual cenário da abertura de dados governamentais, seja no contexto brasileiro ou em nível mundial. Portanto, ao mesmo tempo em que se percebe um avanço na investigação da disponibilização dos dados abertos governamentais para a sociedade, nota-se a ausência de avaliação dos temas, ou seja, quais informações são disponibilizadas por meio de dados abertos nos Institutos Federais.

MÉTODO

Esta seção apresenta uma contextualização do instrumento de análise concebido como fonte de dados para o artigo, assim como as técnicas empregadas para a realização das atividades propostas e as características do modelo de análise criado. É importante ressaltar que o levantamento bibliográfico de toda a base teórica utilizada neste estudo foi realizado na plataforma *Google Scholar* com pesquisas relacionadas ao tema central em questão, sendo analisados artigos científicos, livros, teses e dissertações. Os artefatos de legislação foram obtidos por meio de acessos a repositórios governamentais.

Contextualização do instrumento de análise

Para elaborar um PDA, as instituições públicas devem realizar discussões com as áreas finalísticas do órgão, definir os responsáveis pela sua criação e implementação, listar todas as bases de dados, permitir que a sociedade informe quais bases de dados serão abertas e, por fim, elaborar uma matriz de priorização informando as bases disponibilizadas (CGU, 2020).

O Comitê Gestor da Infraestrutura Nacional de Dados Abertos (CGINDA) é formado por membros convidados da sociedade civil, setor acadêmico, Ministérios, Secretaria Geral da Presidência da República, Casa Civil da República e do Instituto Brasileiro de Geografia e Estatística (IBGE) Aleixo (2020). A priorização da abertura de bases de dados demandadas pela sociedade é evidenciada na seguinte resolução do CGINDA:

Resolução nº 3, de 13 de outubro de 2017 (2017)

Art. 1º Para promover a cultura de transparência pública, conforme inciso II do § 2º do art. 5º do Decreto nº 8.777, de 11 de maio de 2016, as bases de dados a serem disponibilizadas devem ser priorizadas e justificadas, nos Planos de Dados Abertos - PDA, em função de seu potencial em termos de interesse público, considerando-se o que for aplicável:

I - o grau de relevância para o cidadão;

II - o estímulo ao controle social;

III - a obrigatoriedade legal ou compromisso assumido de disponibilização daquele dado;

IV - o dado se referir a projetos estratégicos do governo;

V - o dado demonstrar resultados diretos e efetivos dos serviços públicos disponibilizados ao cidadão pelo Estado;

VI - a sua capacidade de fomento ao desenvolvimento sustentável;

VII - a possibilidade de fomento a negócios na sociedade;

VIII - os dados mais solicitados em transparência passiva desde o início da vigência da Lei nº 12.527, de 18 de novembro de 2011 - Lei de Acesso à Informação, (Resolução nº 3, 2017, s/p).

Com o intuito de auxiliar o processo de publicação de bases de dados, a CGU sugere uma matriz de priorização para conjuntos de dados de Universidades e Institutos Federais baseada nos dados que são comumente solicitados

pela sociedade civil, perfazendo um total de 77 informações (itens) em 14 temas. Todos esses itens e temas são apresentados na Tabela 1, a qual dispõe as informações constantes em CGU (2020) de forma adaptada graficamente.

Nº	Tema	Informações
1	Alunos	Quantidade de alunos, Quantidade de alunos por cursos, Quantidade de abandonos, Matrículas trancadas, Quantidade de alunos formados
2	Cursos	Nome, Coordenador, E-mail, Telefone, Modalidade (Presencial ou EAD) Grau (licenciatura ou bacharelado), UF, Município, Vagas Autorizadas, Duração, Turno
3	Disciplinas	Nome da disciplina, Carga Horária, Quantidade de Vagas, Pré-requisitos, Cursos relacionados
4	Professores por Departamento	Nome do professor, Nome do departamento, Carga Horária, Linha de Pesquisa
5	Horário de atividades semanais dos docentes	Nome do docente, Nome do departamento, Horários, Estatísticas
6	Estatísticas	Média do IRA (Índice de Rendimento Acadêmico), Índice de reprovação por matéria, Percentual de Preenchimento das Cotas, Patentes (data depósito, número do pedido, título, inventores, tema), atendimentos por tipo de serviço, atividade de extensão, acervo da biblioteca, lista de postos de atendimento, equipamentos públicos
7	Assistência ao estudante	Tipo do auxílio, Gastos com o programa, Quantidade de beneficiados
8	Banco de TCCs, monografias, artigos e teses	Título, Resumo, Autores, Palavras-chave, Data do documento
9	Produção científica	Título da produção, Resumo, Autores, Palavras-chave, Data do documento, URL, Periódico, Conferência
10	Projetos que envolvem recebimento de recursos externos	Acordos de Cooperação Técnica (Nacionais e Internacionais), Termos de Execução Descentralizada, Intercâmbio, Convênios fora do SICONV
11	Empresas incubadas	Quantidade de empresas incubadas, Finalidades, áreas de atuação, Lista de empresas juniores, Lista de empregados terceirizados
12	Obras	Obras em andamento, Conclusão de Obras, Custo por obra
13	Concursos públicos	Cargos, Quantidade de vagas, Classe, Código SIAPE, Código da vaga
14	Lista de imóveis para alugar	Tipo do imóvel, Localização, Destinação, Quantidade de quartos, Valor do aluguel, Valor do condomínio, Valor do imóvel, Situação(alugado/vago)

Tabela 1. Matriz de priorização de bases de dados sugerida pela CGU

Fonte: Adaptado de CGU (2020).

Coleta de dados e modelo de análise

Este trabalho realizou uma pesquisa quantitativa por meio da técnica de pesquisa documental no [Portal Brasileiro de Dados Abertos \(PBDA\)](#), analisando os conjuntos de dados disponibilizados pelos Institutos Federais em relação aos temas sugeridos pela CGU para abertura de bases de dados. Apesar de constatado que seis Institutos Federais disponibilizaram mais conjuntos de dados no seu próprio site institucional, apenas foram analisados os conjuntos de dados catalogados no PBDA, pois somente “serão consideradas como bases de dados efetivamente publicadas as que estiverem catalogadas no referido Portal” (CGU, 2020, p. 10).

Entre os dias 15 de abril a 01 de maio de 2021, foram analisados 543 conjuntos de dados abertos de 38 Institutos Federais disponíveis no PBDA. No intuito de verificar se os temas e informações sugeridas pela CGU estão sendo disponibilizados por meio de dados abertos por cada um dos Institutos Federais, foram analisados o título, a descrição, o dicionário de dados e o conteúdo de cada um dos conjuntos, os quais estão disponíveis em <https://tinyurl.com/ogdif2021>.

Para analisar os 77 itens (informações) de cada um dos 14 temas, estabeleceu-se um índice na forma de uma escala somada de 3 pontos, visando a mensurar o grau de transparência da informação solicitada, sendo atribuído “0” (zero) para o item com disponibilização de dados inexistente, “1” (um) para o item com disponibilização parcial de dados e “2” (dois) para o item com disponibilização completa de dados. Assim, criou-se um indicador de comparação dos 14 temas, considerando a pontuação obtida a partir da soma dos valores de seus respectivos itens. A interpretação dos itens por meio de índices permitiu, portanto, a análise parametrizada de cada um dos temas. Para o cálculo do indicador de cada tema, considerou-se a pontuação média por item ajustada para um escore de “0” a “10”, onde “0” indica que nenhum dos itens do respectivo tema teve seus dados disponibilizados,

mesmo que parcialmente, e “10” indica que todos os itens do respectivo tema tiveram seus dados integralmente disponibilizados.

RESULTADOS

Após o processo de coleta de dados, identificou-se que 73,70% dos Institutos Federais catalogaram pelo menos uma base de dados no Portal Brasileiro de Dados Abertos. Desses, 60,71% não descreveram seus conjuntos de dados por meio de dicionários de dados. Adicionalmente, alguns *links* de conjuntos de dados apresentaram “erro 404”, isto é, a página que deveria conter o arquivo não foi encontrada. Por consequência disso, esses conjuntos de dados não foram analisados. Um número considerável de conjuntos de dados não apresentava descrições, sendo assim, a análise do tema foi limitada ao título e conteúdo dos dados em seus respectivos arquivos. Nenhum Instituto Federal disponibilizou ou referenciou ontologias em seus conjuntos de dados abertos. Por fim, diversos conjuntos de dados estavam desatualizados ou não continham qualquer data referente às informações disponibilizadas; apesar disso, a análise acerca do tema foi realizada, uma vez que a análise temporal não faz parte do escopo desta pesquisa. A Tabela 2 apresenta os resultados dos 14 temas consolidados por Instituto Federal.

Instituto	1	2	3	4	5	6	7	8	9	10	11	12	13	14
IFAC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFAL	0	5	0	5	0	2,22	10	0	0	5	0	0	8	0
IFAM	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFAP	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFB	10	0	0	8,75	0	3,33	0	5	5	7,5	0	0	0	0
IFBA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFBAIANO	2	0	6	5	0	1,11	0	0	0	0	0	0	0	0
IFC	4	0	0	7,5	0	2,2	10	0	0	5	0	0	0	0
IFCE	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFES	2	3,64	0	7,5	0	1,11	6,67	0	0	0	0	0	4	0
IFF	0	3,18	0	7,5	0	0	0	0	0	0	0	0	8	0
IFFARROUPILHA	4	4,09	4	5	0	1,11	0	0	0	0	0	0	0	0
IFG	8	6,82	5	5	0	1,67	0	0	0	0	0	1,67	6	0
IFGOIANO	1	3,18	0	8,75	0	0	0	0	0	0	0	0	0	0
IFMA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFMG	4	4,09	0	5	0	2,22	0	0	0	2,5	0	10	4	0
IFMT	10	3,18	0	8,75	0	2,22	0	0	0	0	0	1,67	0	0
IFMS	8	4,09	0	8,75	0	1,11	0	0	0	0	8,75	10	6	0
IFNMG	8	6,82	0	2,5	0	1,11	10	0	0	2,5	0	0	4	0
IFPA	4	0	0	1,25	0	3,33	5	0	0	5	0	0	5	0
IFPB	10	4,55	2	7,5	0	1,11	6,67	0	0	0	0	0	0	0
IFPE	0	0	0	0	0	0	0	0	1,25	1,25	0	0	0	0
IFPI	0	0	0	5	0	0	0	0	0	0	0	0	0	0
IFPR	10	5	3	5	0	2,22	10	0	0	5	0	0	2	0
IFRJ	2	5,91	0	0	0	1,11	0	0	0	0	0	0	0	0
IFRN	10	0,91	0	7,5	8,33	2,22	3,33	0	0	0	0	0	0	0
IFRO	0	5,91	3	5	0	1,11	0	0	0	2,5	0	1,67	0	0
IFRR	0	0	0	7,5	0	0	0	0	0	0	0	0	0	0
IFRS	0	0	0	0	0	1,11	0	0	0	0	0	0	4	0
IFS	8	5	4	7,5	0	2,22	0	0	0	2,5	0	6,67	4	0
IFSC	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFSP	8	5	0	7,5	0	2,22	0	0	6,25	2,5	0	0	0	0
IFSPE	10	6,82	0	7,5	0	1,11	0	0	0	0	0	0	0	0
IFSUDESTEMG	10	5	6	7,5	0	1,67	6,67	0	0	0	0	0	0	0
IFSUL	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IFSULDEMINAS	4	5,45	4	7,5	0	1,11	6,67	0	0	0	0	10	0	0
IFTM	8	4,09	4	7,5	0	3,33	6,67	0	6,25	0	0	6	0	0
IFTO	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabela 2. Escore de disponibilização por temas sugeridos pela CGU

Nota: 1 - Alunos; 2 - Cursos; 3 - Disciplinas; 4 - Professores por Departamento; 5 - Horário de atividades semanais dos docentes; 6 - Estatísticas; 7 - Assistência ao estudante; 8 - Banco de TCCs, monografias, artigos e teses; 9 - Produção científica; 10 - Projetos que envolvem o recebimento de recursos externos; 11 - Empresas incubadas; 12 - Obras; 13 - Concursos Públicos; 14 - Lista de imóveis para aluguel.

A média dos escores (dentro do interstício de 0 a 10) dos temas por região foram as seguintes: região Centro-Oeste

2,27 (desvio padrão: 0,93); região Sudeste 2,16 (desvio padrão: 0,80); região Nordeste 1,21 (desvio padrão: 1,15); região Sul 1,12 (desvio padrão: 1,23); região Norte 0,51 (desvio padrão: 0,73).

Nenhum dos itens foi disponibilizado por todos os 38 Institutos Federais. Especificamente sobre os itens pertinentes ao tema “Imóveis para alugar”, todos os Institutos Federais negligenciaram a apresentação dos dados. Ademais, os itens “e-mail” e “telefone” do tema “cursos”, os itens “quantidade de vagas” do tema “disciplinas”, o item “acervo de biblioteca” do tema “estatísticas” e o item “palavras-chave” do tema “produção científica” não foram disponibilizados por nenhum Instituto Federal. Os itens “Nome do professor” e “Nome do departamento” do tema “Professores por Departamento”, os itens “Nome” e “Grau (licenciatura ou bacharelado)” do tema “Cursos” e o item “atividade de extensão” do tema “Estatísticas” foram disponibilizados por, pelo menos, a metade dos Institutos Federais.

Conforme observado na Figura 1, os temas “4 (Professores por Departamento)” e “1 (Alunos)” apresentaram os melhores resultados em termos de disponibilização de dados, porém ainda obtiveram margens inferiores à 50% de todos os Institutos Federais.

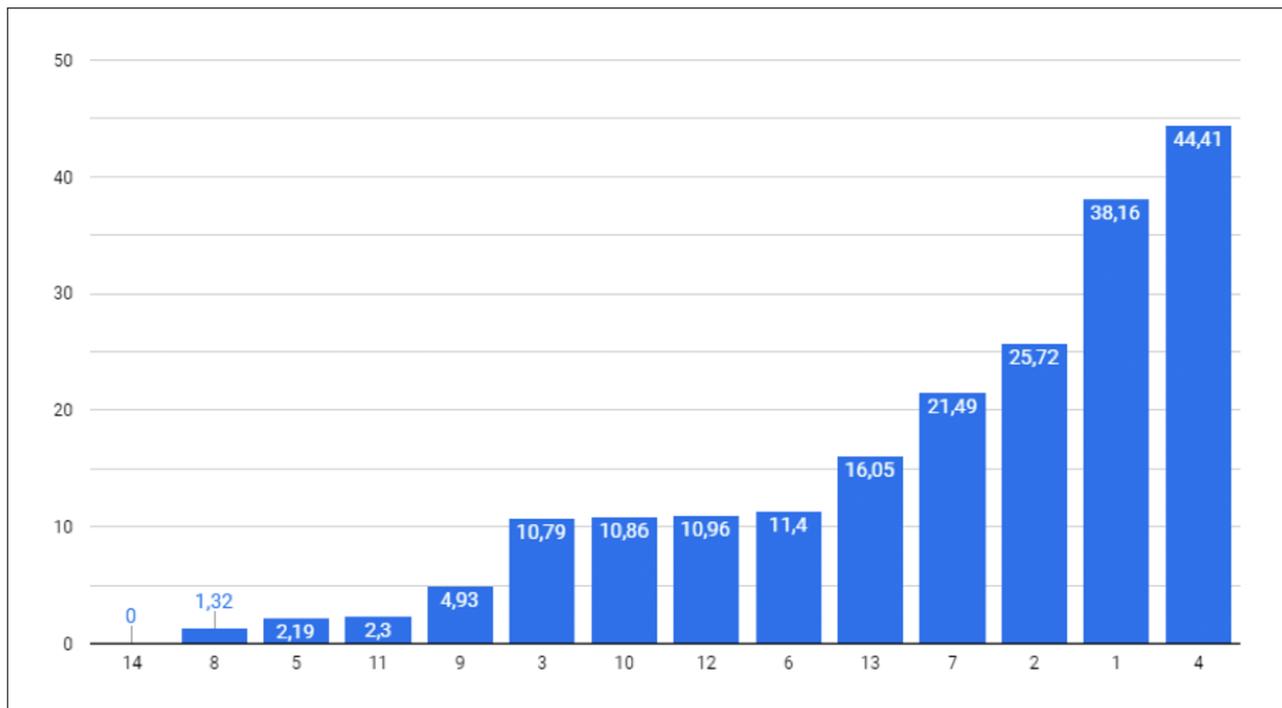


Figura 1. Porcentagem de disponibilização dos temas sugeridos pela CGU.

Nota: 1 - Alunos; 2 - Cursos; 3 - Disciplinas; 4 - Professores por Departamento; 5 - Horário de atividades semanais dos docentes; 6 - Estatísticas; 7 - Assistência ao estudante; 8 - Banco de TCCs, monografias, artigos e teses; 9 - Produção científica; 10 - Projetos que envolvem o recebimento de recursos externos; 11 - Empresas incubadas; 12 - Obras; 13 - Concursos Públicos; 14 - Lista de imóveis para alugar.

Discussão

Os resultados desse estudo indicam baixa disponibilização de dados relativos aos temas e itens sugeridos pela CGU. A inexistência de qualquer informação catalogada por alguns Institutos Federais, assim como a ausência de dicionários de dados e descrição dos conjuntos de dados dificultaram a análise dos dados em relação aos temas e itens. Esses resultados estão em consonância com a afirmação de Klein (2018) Klein (2018) de que nenhum Instituto Federal é 100% transparente, considerando-se a disponibilização de seus dados conforme preconizado pela CGU. Adicionalmente, a baixa publicação de dados e ausência de organização das informações impactam negativamente no consumo e reuso dos dados abertos governamentais, conforme evidenciado em outros estudos correlatos (p. ex. Lima et al., 2017; Silva, 2018).

Neste sentido, a ausência do emprego de ontologias nos dados abertos dos Institutos Federais pode dificultar a interoperabilidade e recuperação destes em um ambiente semântico, tal qual evidenciado por Santarem Segundo (2015). Ademais, uma das causas da baixa disponibilização de determinados dados abertos provavelmente está relacionada à ausência de sistemas de Tecnologia da Informação que facilitem a sua extração e oferta na Web Nunes (2018).

A alta variabilidade dos indicadores evidencia a discrepância entre as instituições, inclusive dentro de uma mesma região. Ou seja, enquanto um Instituto Federal disponibilizou informações sobre diversos itens, outro não divulgou quaisquer informações. Constatou-se, ainda, a falta de sincronia entre a publicação de dados abertos no próprio site institucional e PBDA. Essa realidade de falta de maturidade e padronização dos Institutos Federais no processo de abertura e manutenção dos seus dados abertos foi observado por Lemos (2017).

É possível agrupar os três temas mais disponibilizados (“Professores por Departamento”, “Cursos” e “Alunos”) como categoria “Ensino”, e alguns dos temas menos disponibilizados (“Produção Científica”, “Banco de TCCs, monografias, artigos e teses”) como categoria “Pesquisa”, significando a ausência de sistemas de Tecnologia da Informação integrados nos Institutos Federais. Em outras palavras, utiliza-se um sistema para gerenciar os dados do ensino, o qual dispõe de funcionalidades para a abertura dos dados, e outro sistema - sem esta capacidade - para controlar os dados da área de pesquisa.

Por fim, os itens do tema “Imóveis para alugar” não foram disponibilizados por nenhum Instituto Federal. Considerando que a CGU sugeriu os mesmos temas para as Universidades e Institutos Federais, acreditamos que dados sobre este tema sejam mais frequentes para as Universidades Federais, devido ao porte e estrutura dessas instituições.

CONCLUSÕES

Este artigo analisou os conjuntos de dados abertos dos Institutos Federais de Educação, Ciência e Tecnologia, tendo como base as sugestões constantes no Manual de Elaboração de Planos de Dados Abertos (PDAs) da CGU (2020), acerca de temas de bases de dados que poderiam ser abertas para a sociedade. Tais sugestões corroboram com o interesse público para a melhor transparência e governança das instituições públicas.

Considerando o objetivo proposto pelo estudo, verificou-se contundentemente diversas lacunas na disponibilização de dados abertos por parte dos Institutos Federais, além da baixa aderência à abertura de dados em conformidade às recomendações da CGU. Encontrou-se, ainda, considerável variação no que tange à publicação dos dados por instituição, além da ausência de sincronia entre o que é apresentado no PBDA e nos repositórios institucionais. Em termos funcionais, cabe compreender a relevância da oferta das informações aqui tratadas para a gestão pública, analisando-se, prioritariamente, os meios administrativos e tecnológicos para atender a essas recomendações.

Uma vez que este estudo indica uma baixa disponibilização dos dados acerca dos temas sugeridos pela CGU, cabe destacar, enquanto limitação da pesquisa, que não foi verificado se os Institutos Federais possuem estes dados para serem publicados de outra forma, ou se a informação em específico não se aplicava ao contexto de cada instituição. Por exemplo, o tema “empresas incubadas” foi disponibilizado por apenas 2,3% dos Institutos Federais, no entanto, caso uma instituição ainda não tenha iniciado nenhum processo de incubação de empresas, seja por falta de demanda, seja por indisponibilidade de infraestrutura ou marco legal, significa que a mesma não possui os dados deste tema para disponibilizar à sociedade.

Com o intuito de continuar a investigação acerca da abertura de dados abertos por Institutos Federais, pesquisas complementares são necessárias para: (1) compreender os motivos da baixa disponibilização de dados abertos por alguns Institutos Federais; (2) analisar e categorizar todos os temas que os Institutos Federais identificaram por meio de audiências públicas e informaram em seus respectivos PDAs; e (3) analisar e categorizar quais temas e informações estão disponíveis por meio de dados abertos dos Institutos Federais, isto é, diferentemente do escopo desta pesquisa, considerar temas que não foram sugeridos pela CGU.

REFERÊNCIAS

- Aleixo, D. V. B. S. (2020). O estado de anomia dos dados no acesso aos dados governamentais abertos no Brasil [Tese de doutorado]. *Instituto Universitário de Lisboa*. Recuperado de <https://repositorio.unesp.br/handle/11449/191686>
- Alencar, A., Xavier, D., Chaves, L., & Souza, D. (2018). Publicação e consumo de dados abertos conecta-dos acadêmicos. *Principia*, 42(2018), 136-145. Recuperado de <https://periodicos.ifpb.edu.br/index.php/principia/article/view/1988>
- Bergeron, P. (1996). Information resources management. *Annual Review of Information Science and Technology*, 31(1996), 263-300.
- Breitman, K. (2005). *Web semântica: a internet do futuro*. Curitiba: LTC.
- Britto, G. C., Ruy, F. B., & Azevedo, C. L. B. (2020). Um ambiente para integração de dados abertos relativos à despesa pública. *Ontobras*, 13(2020). Recuperado de <http://ceur-ws.org/Vol-2728/paper13.pdf>
- Cardoso, P. H. (2019). Ciência de dados aplicada a dados governamentais abertos sob a ótica da Ciência da Informação [Dissertação de mestrado]. *Instituto Universitário de Lisboa*. Recuperado de <https://ridi.ibict.br/handle/123456789/1039>
- Carossi, D. F., Pedan, J. M., & Teixeira Filho, J. G. (2020). Utilização da metodologia "Grounded Theory" na identificação de informações para compor um plano de dados abertos de instituições federais de ensino superior. *iSys-Revista Brasileira de Sistemas de Informação*, 10(2), 171-192. Recuperado de <http://seer.unirio.br/index.php/isys/article/view/6152>
- Chen, C., Huang, H., Ross, K. E., Cowart, J. E., Arighi, C. N., Wu, C. H., & Natale, D. A. (2020). Protein ontology on the semantic web for knowledge discovery. *Scientific Data*, 7(337), 1-12. Recuperado de <https://www.nature.com/articles/s41597-020-00679-9>
- Choo, C. W. (1996). The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International journal of information management*, 16(5), 329-340. Recuperado de <https://www.sciencedirect.com/science/article/abs/pii/0268401296000205>
- Controladoria-Geral da União [CGU]. (2020, jun). *Manual de elaboração de planos de dados abertos (PDAs)*. Brasília, DF: STPC/CGU. Recuperado de <https://repositorio.cgu.gov.br/handle/1/46702>
- Gonçalves, V. (2017). *PSO: uma ontologia de domínio sobre servidores públicos federais* (Dissertação de mestrado, Universidade Federal de Pernambuco, Brasil). Recuperado de <https://repositorio.ufpe.br/handle/123456789/29774?mode=full>
- Klein, C. R. (2018). *Transparência da gestão pública das autarquias federais de ensino vinculadas ao Ministério da Educação - MEC* (Dissertação de mestrado, Universidade de Brasília, Brasil). Recuperado de <https://repositorio.unb.br/handle/10482/32187>
- Lemos, J. M. M. (2017). *Um processo para publicação de dados abertos em institutos federais baseado em BPM* (Dissertação de mestrado, Universidade Federal de Pernambuco, Brasil). Recuperado de <https://repositorio.ufpe.br/handle/123456789/25351>
- Lima, J. P., Cunha, S. F., & Maior Filho, U. G. S. (2017). *Governo Aberto: um estudo de caso sobre o Instituto Federal da Paraíba*. João Pessoa, PB, Brasil: Encontro Brasileiro de Administração Pública. Recuperado de <http://plone.ufpb.br/ebap/contents/documentos/0073-088-governo-aberto-um-estudo-de-caso-sobre-o-ifpb.pdf>
- Macedo, D. F., & Lemos, D. L. d. S. (2021, abr.). Dados abertos governamentais: iniciativas e desafios na abertura de dados no brasil e outras esferas internacionais. *AtoZ: novas práticas em informação e conhecimento*, 10(2), 14. Recuperado de <https://doi.org/10.5380/atoz.v10i2.77737> doi: 10.5380/atoz.v10i2.77737
- Monteiro, G., & Lucas, E. R. d. O. (2019, jun.). Dados científicos abertos: identificando o papel das políticas de gestão e das agências de fomento. *AtoZ: novas práticas em informação e conhecimento*, 8(1), 13. Recuperado de <https://doi.org/10.5380/atoz.v8i1.67253> doi: 10.5380/atoz.v8i1.67253
- Nunes, V. K. A. (2018). *Plano de dados abertos nos institutos federais de Educação, Ciência e Tecnologia* (Dissertação de mestrado, Universidade Federal de Viçosa, Brasil). Recuperado de <https://www.locus.ufv.br/handle/123456789/22591>
- Possamai, A. J. (2016). *Dados abertos no governo federal brasileiro: desafios de transparência e interoperabilidade* (Tese de doutorado, Universidade Federal do Rio Grande do Sul, Brasil). Recuperado de <https://www.lume.ufrgs.br/handle/10183/156363>
- Resolução nº 3, de 13 de outubro de 2017. (2017). *Approva as normas sobre elaboração e publicação de Planos de Dados Abertos, conforme disposto no Decreto nº 8.777, de 11 de maio de 2016*. Brasília, DF. Recuperado de <http://plone.ufpb.br/ebap/contents/documentos/0073-088-governo-aberto-um-estudo-de-caso-sobre-o-ifpb.pdf>
- Santarem Segundo, J. E. (2015). Web semântica, dados ligados e dados abertos: uma visão dos desafios do brasil frente às iniciativas internacionais. *Tendências da Pesquisa Brasileira em Ciência da Informação*, 8(2), 219-239. Recuperado de <https://revistas.ancib.org/index.php/tpbci/article/view/359>
- Silva, P. N. (2018). *Dados Governamentais Abertos: Métricas e indicadores de Reuso* (Tese de doutorado, Universidade Federal de Minas Gerais, Brasil). Recuperado de <https://repositorio.ufmg.br/handle/1843/BUBD-AYNG4U>
- World Wide Web Consortium [W3C]. (2017). *Data on the web best practices: W3C recommendation 31 January 2017*. Recuperado de <https://www.w3.org/TR/dwbp/>
- Zaidan, F. H., & Bax, M. P. (2013, ago.). Linked open data como forma de agregar valor às informações clínicas. *AtoZ: novas práticas em informação e conhecimento*, 2(1), 44. Recuperado de <https://doi.org/10.5380/atoz.v2i1.41319> doi: 10.5380/atoz.v2i1.41319

Como citar este artigo (APA):

Marques, Y. B., Bax, M. P., Lopes Júnior, G., Souza, G. H. S. & Melo, E. C. (2022). Dados abertos dos Institutos Federais de Educação: análise de temas sugeridos pela Controladoria-Geral da União. *AtoZ: novas práticas em informação e conhecimento*, 11, 1 – 10. Recuperado de: <http://dx.doi.org/10.5380/atoz.v11.82605>

FORMULÁRIO DE NOTAS DA OBRA E CONFORMIDADE COM A CIÊNCIA ABERTA

CONTRIBUIÇÃO DE AUTORIA

Papéis e contribuições	Yuri Bento Marques	Marcello Peixoto Bax	Geraldo Lopes Júnior	Gustavo Henrique Silva de Souza	Eduardo Cardoso Melo
Concepção do manuscrito	X	X	X	X	X
Escrita do manuscrito	X	X	X	X	X
Metodologia	X	X	X	X	
Curadoria dos dados	X	X			
Discussão dos resultados	X	X	X	X	
Análise dos dados	X	X			X

CONFORMIDADE COM A CIÊNCIA ABERTA

Disponibilidade de Dados Científicos da Pesquisa

Os conteúdos subjacentes ao texto da pesquisa estão disponíveis sob demanda dos pareceristas.

EQUIPE EDITORIAL

Editora/Editor Chefe

Maria do Carmo Duarte Freitas (<https://orcid.org/0000-0002-7046-6020>)

Editora/Editor Adjunta/Adjunto

Paula Carina de Araújo (<https://orcid.org/0000-0003-4608-752X>)

Editora/Editor de Texto Responsável

Fabiane Führ (<https://orcid.org/0000-0002-3723-050X>)

Nicholle Ferreira Murmel Liali (<https://orcid.org/0000-0002-1086-908X>)

Editora/Editor de Layout

André José Ribeiro Guimarães (<https://orcid.org/0000-0003-0874-7400>)