

XXV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXV ENANCIB

GT 8 – Dados, Informação e Tecnologia

DADOS PÚBLICOS SOBRE A EDUCAÇÃO DE JOVENS E ADULTOS

PUBLIC DATA ABOUT YOUTH AND ADULT EDUCATION

Gabriela Melo Rocha - Universidade de Brasília (UnB)
Michelli Pereira da Costa - Universidade de Brasília (UnB)

Modalidade: Resumo Expandido

Resumo: discute o uso de dados governamentais abertos como instrumento para análise crítica e transformação social, com foco na Educação de Jovens e Adultos (EJA) no Brasil. O projeto piloto aqui apresentado avaliou a viabilidade técnica das aplicações adotadas para sistematização e visualização de dados públicos sobre educação no Portal dos Fóruns de EJA. O método utilizado para desenvolver o protótipo considerou cinco dos sete estágios da visualização de Fry (2007), abarcando os dados de matrícula da EJA entre 2013 e 2024 e disponibilizando-os na plataforma Visão. Essa iniciativa mostra o potencial de uso de dados governamentais abertos em contextos tecnopolíticos.

Palavras-chave: educação de jovens e adultos; dados governamentais abertos; visualização de dados.

Abstract: this work discusses the use of open government data as a tool for critical analysis and social transformation, focusing on Youth and Adult Education (EJA) in Brazil. The pilot project presented here evaluated the technical feasibility of the applications adopted for the systematization and visualization of public education data on the EJA Forums Portal. The method used to develop the prototype considered five of Fry's seven stages of visualization (2007), encompassing EJA enrollment data between 2013 and 2024 and making it available on the Visão platform. This initiative demonstrates the potential for using open government data in technopolitical contexts.

Keywords: youth and adult education; open government data; data visualization.

1 INTRODUÇÃO

Na atual organização e produção do conhecimento, os dados são essenciais para a descrição de fenômenos, realidades e objetos, compondo parte da informação a ser produzida, transmitida e utilizada. Como produtos da comunicação e da atividade humana, os dados refletem os interesses dos grupos que influenciam sua produção e interpretação.

Nesse contexto, a abertura de dados governamentais, embora ainda não plenamente adequada, representou uma importante conquista dos movimentos sociais e da participação cidadã. No entanto, persistem obstáculos para a sua reutilização, especialmente quando o interessado possui pouca familiaridade com ferramentas tecnológicas de coleta e análise. Esse

cenário se aplica ao uso de dados públicos no contexto dos Fóruns de EJA Brasil, movimento social em defesa da educação de jovens, adultos e trabalhadores que busca utilizar essas informações para fortalecer o debate, embasar demandas e subsidiar pesquisas relacionadas às pautas que estruturam o movimento.

Assim, este estudo teve como problema a seguinte questão: Como aplicar estratégias de tratamento e disponibilização de dados públicos sobre educação no contexto das articulações tecnopolíticas dos Fóruns de EJA Brasil? O principal objetivo, é discutir e propor um produto de visualização da taxa de matrícula na EJA, considerando as etapas de coleta, normalização e disponibilização. Nesse sentido, o estudo proposto apresenta potencial para contribuir teoricamente com as reflexões sobre a publicização e visualização de dados educacionais, ao mesmo tempo em que oferece aportes práticos para o desenvolvimento do projeto em curso no âmbito do Portal dos Fóruns de EJA Brasil. Trata-se, portanto, de uma iniciativa que articula pesquisa e ação, fortalecendo tanto a produção de conhecimento quanto sua aplicação em contextos concretos de agência social.

2 DADOS GOVERNAMENTAIS ABERTOS E DADOS PÚBLICOS SOBRE EDUCAÇÃO

No âmbito das políticas governamentais, os dados adquiriram centralidade como instrumentos de gestão pública e como objeto de disputa política. Nesse cenário, intensificaram-se as reivindicações sociais por maior controle e transparência na atuação do Estado, especialmente no que diz respeito à formulação e implementação de políticas públicas. No Brasil, esse debate ganhou força com as mobilizações em torno das iniciativas relacionadas à agenda do governo aberto (Brasil, 2011; CGU, 2014). Tais movimentos visam ampliar a participação social na gestão pública, assegurando a disponibilização aberta de informações, particularmente de dados públicos, como estratégia de fortalecimento do acompanhamento social e da tomada de decisões coletivas.

A proposta de abertura dos dados governamentais é resultado das conquistas dos movimentos sociais organizados, que reivindicam maior acesso e transparência às informações produzidas pelo poder público. Oliveira e Silva (2024) destacam o movimento em prol do acesso aberto às publicações científicas, iniciado nos anos 2000, como precursor dos princípios que sustentam essas reivindicações. A insatisfação com modelos tradicionais de comunicação, aliados às novas possibilidades de compartilhamento propiciadas pelas

Tecnologias da Informação e Comunicação (TICs), impulsionou a defesa da abertura dos dados.

A discussão sobre dados governamentais abertos esteve vinculada às ações da Open Government Partnership (OGP), organização internacional que promove a agenda de governo aberto e a cultura da transparência. No Brasil, destaca-se a aprovação da Lei de Acesso à Informação (LAI) e a consolidação dos princípios de transparência, prestação de contas e participação cidadã. Em conformidade com o princípio da transparência, informações e dados governamentais de interesse público devem ser acessíveis a toda a sociedade. Para tanto, o governo brasileiro adota estratégias de transparência ativa, a partir da disponibilização espontânea de dados e de transparência passiva, por meio da divulgação mediante solicitação da sociedade (Braga; Costa, 2021).

De acordo com o Open Data Institute (ODI), os dados podem ser classificados como pessoais, comerciais ou governamentais. Este estudo restringe-se à análise dos dados governamentais abertos, ou seja, dados produzidos por órgãos do governo e compartilhados com o público sob uma licença aberta de distribuição. Essa licença deve garantir mais do que a simples disponibilidade dos dados; deve assegurar que eles sejam acessíveis sem necessidade de login específico, legíveis por máquinas e passíveis de compartilhamento e reutilização em diversos contextos

Visando orientar a abertura qualificada dos dados, Bizer, Heath e Berners-Lee (2009), propuseram a utilização de um sistema de classificação em cinco estrelas para dados abertos. A classificação estabelece parâmetros para diferentes níveis de abertura, nos quais o primeiro nível representa o estágio mais básico e o quinto o nível mais desejável, no qual os dados podem atingir seu máximo potencial de reutilização.

No contexto dos dados governamentais Wang e Shepherd (2020) observam que a maioria dos conjuntos de dados disponibilizados pelo governo do Reino Unido apresentavam sérios problemas na estrutura da abertura, sendo classificados com apenas uma estrela segundo o modelo proposto por Bizer, Heath e Berners-Lee (2009). Esse cenário, também presente no Brasil, representa um obstáculo significativo para a promoção da participação cidadã e o fortalecimento do engajamento social a partir do uso de informações públicas.

Com vistas a superar parte desses desafios, diversos sistemas de visualização e manipulação de dados públicos foram desenvolvidos ao longo do tempo, configurando-se como alternativas para ampliar o acesso, a compreensão e a apropriação dos dados pela

sociedade. Foi a partir dessa problemática que o Portal dos Fóruns de Educação de Jovens e Adultos (EJA)¹ se mobilizou para coletar, tratar e disponibilizar dados públicos sobre educação, constituindo uma experiência que será discutida neste estudo, com atenção especial às técnicas e tecnologias adotadas.

O Portal dos Fóruns de EJA foi criado em 2005, fundamentado nos princípios do movimento de *software livre*. O acervo do Portal é alimentado de forma descentralizada por representantes dos fóruns estaduais brasileiros e é composto majoritariamente por publicações (Cunha, 2017). Além do acervo de mais de 17 mil publicações, o Portal disponibiliza e promove a visualizações de dados públicos sobre educação, compreendendo que esses dados constituem ferramenta fundamental para o diálogo com o poder público e para o fortalecimento das políticas educacionais.

3 METODOLOGIA

Para viabilizar os objetivos da pesquisa de discutir e propor um produto de visualização de dados públicos sobre a EJA, foram adotadas cinco etapas das sete sugeridas por Radburn, Dykes e Wood (2010), baseadas no modelo de Fry (2004; 2007). Tal proposta metodológica compreende sete estágios de desenvolvimento, que vão desde a coleta dos dados até a interação do usuário com os recursos visuais. Os sete estágios do modelo de Fry são: 1. Adquirir; 2. Processar; 3. Filtrar, 4. Minerar; 5. Representar; 6. Refinar e 7. Interagir. No estudo do protótipo aqui apresentado, foram considerados apenas os cinco primeiros estágios. A exclusão dos dois estágios finais deve-se ao fato de que na fase atual do projeto, ainda não é possível realizar estudo e de interação do usuário com o sistema desenvolvido.

Os estágios considerados neste estudo foram reorganizados em três fases de atuação (Figura 1) com os métodos correspondentes do modelo de Fry (2007). Essa reorganização justifica-se pela necessidade de agrupar os estágios de acordo com os objetivos e atividades específicas desempenhadas em cada momento do projeto, otimizando os processos e técnicas empregadas. Para cada fase, foram utilizados diferentes plataformas e *softwares*, selecionados conforme a natureza das tarefas e técnica necessária.

¹ Disponível em: <https://forumeja.org.br/>

Figura 1 - Estágios de tratamento dos dados públicos sobre a EJA no contexto do Fórum EJA

| Estágios | Objetivo | Descrição | Ferramentas/ Plataformas |
|---------------------------------|---|--|--|
| 1. Adquirir | Coletar dados públicos sobre a EJA | Identificação e seleção das fontes de dados e coleta de dados | Portal Brasileiro de Dados Abertos, Portal do MEC e página do Inep |
| 2. Processar, filtrar e minerar | Preparar os dados para serem reutilizados em ferramenta de visualização | Seleção de dados e adequação à ferramenta de visualização | Open Refine, Google Planilhas |
| 3. Representar | Carregar os dados em uma ferramenta de visualização | Geração de uma informação visual interativa e aberta à sociedade | Visão |

Fonte: elaborada pelas autoras, adaptado de Fry (2007)

4 TRATAMENTO E VISUALIZAÇÃO DOS DADOS PÚBLICOS SOBRE A EJA NO CONTEXTO DO FÓRUM EJA

No âmbito do Portal dos Fóruns de EJA, a primeira demanda identificada em relação aos dados governamentais abertos concentrou-se nas informações sobre matrícula na EJA em nível nacional. Dessa forma, o desenvolvimento do protótipo apresentado neste trabalho teve como foco avaliar a viabilidade técnica das aplicações adotadas nesse contexto.

4.1 Adquirir

O estágio de aquisição dos dados ocorreu entre os meses de dezembro de 2024 a fevereiro de 2025. A busca por dados educacionais precisou ser realizada diretamente na página do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), uma vez que o repositório nacional de dados abertos governamentais (dados.gov.br) ainda não se encontra devidamente povoado com conjuntos de dados adequados para reutilização. Essa limitação foi evidenciada tanto neste estudo quanto na pesquisa de Penteadó, Bittencourt e Isotani (2017), que apontam fragilidades na estruturação e disponibilização dos dados no portal.

O Inep disponibiliza os dados do Censo Escolar, a principal pesquisa estatística sobre a educação básica no Brasil, todos facilmente acessados na página de resultados do Censo², contém dados de 1995 a 2025 para *download* por meio de planilhas no formato CSV.

²Disponível em: <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar>

Os resultados do Censo Escolar foram sistematizados em dois conjuntos de planilhas:

- Conjunto I – Contém dados sobre ensino regular (redes estaduais e municipais).
- Conjunto II – Contém dados sobre o ensino especial, voltado para estudantes com necessidades educacionais especiais.

O projeto piloto, inicialmente concebido para analisar a série histórica das taxas de matrícula da última década, expandiu seu escopo com a divulgação dos dados de 2024. Assim, a coleta de dados abrangeu 11 anos, de 2013 a 2024, resultando em 22 planilhas: 11 sobre o ensino regular (Conjunto I) e 11 sobre o ensino especial (Conjunto II).

4.2 Processar, filtrar e minerar

Esse estágio, que combina o processamento, filtragem e mineração dos dados, tem por objetivo a normalização e adequação dos dados para reutilização em uma ferramenta de visualização. A escolha da ferramenta visou a adequação com os princípios do *software* livre e a facilidade no manuseio para usuários finais, especialmente porque um dos objetivos do projeto é difundir o uso dos dados públicos para os educandos e educadores da EJA. Tendo em vista esse requisito, foi escolhido o Sistema Aberto de Observatórios para Visualização de Informações (Visão)³, serviço desenvolvido em código aberto pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e destinado ao público em geral.

Além dos parâmetros do Visão, foram utilizadas as ferramentas OpenRefine e Google Planilhas para a limpeza e normalização dos dados. A limpeza visa identificar e corrigir erros e inconsistências, que podem surgir devido a erros de digitação, informações ausentes ou dados inválidos, garantindo assim maior qualidade e confiabilidade das informações (Rahm; Do, 2000).

As planilhas coletadas foram filtradas para incluir apenas os dados referentes à EJA, removendo colunas de quantidade de matrículas parciais e integrais relacionadas à Educação Infantil (creche e pré-escola), Ensino Fundamental (anos iniciais e anos finais) e Ensino Médio. Manteve-se a segmentação EJA Fundamental e EJA Médio. Exclusivamente nos anos de 2013 e 2014 foi necessário somar os dados do ensino EJA parcial e integral, utilizando a fórmula =SOMA() do Google Planilhas⁴.

³ Disponível em: <https://visao.ibict.br/>

⁴ Disponível em: <https://workspace.google.com/intl/pt-BR/products/sheets/>

A normalização dos dados seguiu os parâmetros indicados pelo Visão. A ferramenta utiliza dados qualitativos, quantitativos e geoespaciais para criar um mapa interativo, estruturado em três tipos de informações (Brasil, 2020):

- Indicadores: Números quantitativos representados graficamente no mapa, colorindo regiões para ilustrar estatísticas.
- Filtros: Ferramentas que refinam os indicadores, delimitando áreas no mapa com base em características específicas.
- Camadas: Pontos de localização no mapa, como escolas e hospitais, que podem ser exibidos sozinhos ou combinados com indicadores.

Os dados geoespaciais podem abranger países, estados ou municípios. Para os Fóruns de EJA, foi estabelecido um indicador para representar as taxas de matrícula da EJA em cada município brasileiro. Para iniciar a limpeza e o tratamento dos dados do Inep, foi necessário cruzar a soma total das matrículas – incluindo Estadual Urbana, Estadual Rural, Municipal Urbana, Municipal Rural, Estadual e Municipal – com os seus respectivos municípios. Esse ajuste foi realizado ao posicionar as colunas de valores cinco linhas acima da primeira coluna, denominada Unidades da Federação – Municípios – Dependência Administrativa.

Em seguida, no estágio de mineração, foi elaborada uma planilha contendo os municípios identificados por geocodes, que são códigos geográficos fornecidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE), garantindo a correta associação dos dados ao território nacional.

Para coletar os valores, associá-los ao geocode do município correspondente e obter *insights* informativos, foi utilizada a função PROCV (Procura Vertical) do Google Planilhas, que permite localizar informações em uma tabela e relacionar dados de diferentes colunas. No entanto, durante esse processo, surgiram conflitos de correspondência, pois alguns municípios da lista do IBGE não foram encontrados na base de municípios do Inep devido a diferentes fatores:

- O município de código 2410306 (Serra Caiada) não possui dados disponíveis nas planilhas do Inep, portanto, lhe foi atribuído o valor 0.
- Nas planilhas de 2013 e 2014, tanto do ensino regular quanto do ensino especial, além de Serra Caiada, também não constam os municípios de São Sebastião do Umbuzeiro (2515203), São Vicente do Seridó (2515401) e Brazópolis (3108909), sendo todos registrados com o valor 0 na visualização.

- Como a função realiza uma pesquisa exata, acentos e sinais gráficos nos nomes dos municípios geram erros na correspondência, exigindo sua remoção via OpenRefine.
- Alguns municípios apresentavam grafia incorreta, o que demandou correção utilizando a função `value.replace` do Open Refine⁵.

Os ajustes relatados foram necessários para garantir a padronização dos dados e a correta associação entre as bases do IBGE e do Inep. Ao final desse processo, foram preparadas 44 planilhas para a criação da representação, sendo 22 do ensino regular, divididas entre Ensino Médio e Ensino Fundamental, e 22 do ensino especial, igualmente distribuídas entre essas duas etapas de ensino. Todos os arquivos estão disponíveis na biblioteca do Fórum EJA⁶.

4.3 Representar

Para transformar os dados em informação e torná-los acessíveis à interpretação, elabora-se uma visualização. A visualização de dados, conforme Silva (2019, p. 207) objetiva “comunicar informações ou ideias complexas de forma clara, precisa e eficiente, de uma forma que ajude os usuários a analisar e raciocinar sobre dados e evidências”.

O Visão gera um mapa personalizado, adaptado às categorias de informação selecionadas pelo usuário no momento de sua criação (Brasil, 2020). No mapa desenvolvido neste projeto⁷, é possível interagir com os dados de matrícula de cada município ao longo de 11 anos, facilitando a análise e a compreensão da evolução da EJA no período.

É importante ressaltar que para que os dados subam corretamente no Visão, a coluna com valores gerados com a PROCV deve ser transformada em texto simples, sem os pontos separadores dos milhares.

A possibilidade de adaptar, complementar e compartilhar dados sem a necessidade de conhecimentos avançados em técnicas da Tecnologia da Informação, cria novas utilidades para o portal Fórum EJA, promovendo um uso crítico e colaborativo das informações públicas para fortalecer a luta e a defesa da Educação de Jovens e Adultos.

⁵ Disponível em: <https://openrefine.org/>

⁶ Disponível em: https://forumeja.org.br/colecao-ou-evento/censo-matricula-municipios-2013-2024/?perpage=12&view_mode=masonry&paged=1&order=ASC&orderby=date&fetch_only_meta=&taxquery%5B0%5D%5Btaxonomy%5D=tnc_tax_4791&taxquery%5B0%5D%5Bterms%5D%5B0%5D=57304&taxquery%5B0%5D%5Bcompare%5D=IN&fetch_only=thumbnail%2Ccreation_date%2Ctitle%2Cdescription

⁷ Disponível em: <https://forumeja.org.br/dados-publicos-sobre-a-eja/>

5 CONSIDERAÇÕES FINAIS

O acesso a dados públicos constitui um elemento importante para o monitoramento de políticas e da mobilização social. No contexto da EJA, o uso desses dados permite um entendimento mais amplo da dinâmica das matrículas em face às distribuições por regiões brasileiras e diferentes faixas temporais. A sistematização, organização e disponibilização dessas informações no Portal dos Fóruns de EJA representa a ampliação das possibilidades de uso de recursos informacionais para a instrumentalização da luta EJA e pela emancipação da classe trabalhadora.

O Portal dos Fóruns de EJA se torna, nesse sentido, um espaço fundamental de articulação e instrumentalização das lutas sociais em prol da educação de jovens e adultos. Ao integrar tecnologias de visualização interativa, como o sistema Visão, o portal transcende a simples compilação de informações e apresenta potencial para realizar como uma ferramenta de crítica a articulação social. Dessa forma, ele tem por objetivo contribuir no processo social que visa uma prática educativa crítica e engajada, alinhada aos princípios da educação popular e da justiça social.

REFERÊNCIAS

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data - The Story So Far. **International Journal on Semantic Web and Information Systems (IJSWIS)**, v. 5, n. 3, p. 1–22, jul. 2009.

BRAGA, T. E. N.; COSTA, M. P. DA. Cultura da transparência: uma análise do Poder Executivo Federal. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 21., 2021, Rio de Janeiro. **Anais [...]**. Rio de Janeiro: ANCIB, 2021. Disponível em: <https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/340>. Acesso em: 17 fev. 2025

BRASIL. **Lei no 12.527, de 18 de Novembro de 2011**, 2011. Disponível em: <https://www.lexml.gov.br/urn/urn:lex:br:federal:lei:2011-11-18;12527>. Acesso em: 01 abr. 2025.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. Instituto Brasileiro de Ciência e Tecnologia. Coordenação Geral de Tecnologias de Informação e Informática. **Workshop de apresentação do Visão**. Brasília, 2020. Disponível em: https://visao.ibict.br/wp-content/uploads/2021/07/Tutorial_Visao.pdf. Acesso em: 14 mar. 2025.

CGU. **O que é Governo Aberto**. 2014. Disponível em: <https://www.gov.br/cgu/pt-br/governo-aberto/governo-aberto-no-brasil/o-que-e-governo-aberto>. Acesso em: 01 abr. 2025.

CUNHA, M. C. **Formação política e as tecnologias**: um estudo sobre as possibilidades deste diálogo. Brasília: Editora Paralelo 15, 2017. Disponível em:
http://forumeja.org.br/sites/forumeja.org.br/files/formacao_politica_tecnologias.pdf.
Acesso em: 31 mar. 2025.

FRY, B. J. **Computational Information Design**. 2004. Dissertação (Doutorado em Philosophy in Media Arts and Sciences) - School of Architecture and Planning, Massachusetts Institute of Technology, Cambridge, MA, 2004.

FRY, B. **Visualizing data**. Sebastopol, CA: O'Reilly Media, Inc, 2007.

RADBURN, R.; DYKES, J.; WOOD, J. VizLib: Using the Seven Stages of Visualization to Explore Population Trends and Processes in Local Authority Research. *In*: GIS RESEARCH UK ANNUAL CONFERENCE, 18., London, UK. **Proceedings** [...] London: University College London, 2010.

OLIVEIRA, D. T.; SILVA, P. N. Representação e recuperação de dados governamentais abertos: uma revisão de literatura. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação, v. 22, p. e024029, 13 dez. 2024.

PENTEADO, B. E.; BITTENCOURT, I. I.; ISOTANI, S. Dados abertos educacionais no Brasil e sua preparação para os dados abertos na web. *In*: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (CBIE 2017), 6., 2017. **Anais** [...] São Carlos, SP: [s. n.], 2017. Disponível em:
https://www.researchgate.net/publication/320344888_Dados_abertos_educacionais_no_Brasil_e_sua_preparacao_para_os_dados_abertos_na_web. Acesso em: 01 abr. 2025.

RAHM, E.; DO, H. H. Data cleaning: problems and current approaches. **IEEE Data Engineering Bulletin**, v. 23, n. 4, p. 3-13, University of Leipzig, Alemanha, 2000.

SILVA, F. C. C. da. Visualização de dados: passado, presente e futuro. **Liinc em Revista**, v. 15, n. 2, p. 205-223, Rio de Janeiro, 2019. Disponível em:
<https://doi.org/10.18617/liinc.v15i2.4812> . Acesso em 01 abr. 2025.

WANG, V.; SHEPHERD, D. Exploring the extent of openness of open government data: A critique of open government datasets in the UK. **Government Information Quarterly**, v. 37, n. 1, p. 101405, 1 jan. 2020.