

# Arquitetura de Dados Educacionais como Plataforma para Governo Inteligente - Utilizando dados abertos para apoio à gestão educacional baseada em evidências

Abílio Nogueira Barros<sup>1</sup>, Aldéryck Félix de Albuquerque<sup>2</sup>, Andreza Alencar<sup>1</sup>,  
Rafael Ferreira Mello<sup>1,2</sup>, Gabriel Alves<sup>1</sup>,  
Ibsen Mateus Bittencourt<sup>3</sup>

<sup>1</sup>Departamento de Computação - Universidade Federal Rural de Pernambuco (UFRPE)

<sup>2</sup>Centro de Estudos e Sistemas Avançados do Recife (CESAR)

<sup>3</sup>Universidade Federal de Alagoas (UFAL)

{abilionbarros, derycck}@gmail.com, {andreza.leite, rafael.mello, gabriel.alves}@ufrpe.br,  
ibsen@feac.ufal.br

**Abstract.** *The main objective of the National Education Plan (PNE) is to guide the development of education in the country and its monitoring is characterized as a fundamental activity for educational management. To carry out this monitoring, solutions are needed that provide a reliable data set capable of assisting in evidence-based decision making. Therefore, this article presents an educational data architecture that uses open government data to support intelligent educational management. It provides various public data for monitoring the situation of municipal and state education based on the goals and indicators of the PNE version 2014-2024. This also makes it possible to cross data from different sources to find factors that influence the achievement of such goals. In order to guarantee the construction of reliable evidence, the proposed architecture relies on the use of census databases, such as the Educational Census, coordinated by the National Institute of Educational Studies and Research (INEP). The present study also presents a case study that demonstrates one of the possible applications of this architecture and highlights the potential value of its use for educational management in the country. We therefore conclude that the proposed architecture is a useful instrument to support platforms for evaluating and monitoring educational public policies, contributing to the construction of an Intelligent Government and to promoting improvements in the quality of Brazilian education.*

**Resumo.** *O Plano Nacional de Educação (PNE) tem como objetivo central orientar o desenvolvimento da educação no país e seu monitoramento se caracteriza como uma atividade fundamental para a gestão educacional. Para realizar esse monitoramento se fazem necessárias soluções que forneçam um conjunto de dados confiável e capaz de auxiliar na tomada de decisão baseada em evidências. Diante disso, este artigo apresenta uma arquitetura de dados educacionais que utiliza dados abertos governamentais para apoiar uma gestão educacional inteligente. Esta disponibiliza diversos dados públicos para o acompanhamento da situação da educação municipal e estadual com base nas metas*

*e indicadores do PNE da versão 2014-2024. Esta também possibilita o cruzamento de dados de diversas fontes para encontrar fatores que influenciem no alcance de tais metas. A fim de garantir a construção de evidências confiáveis, a arquitetura proposta conta com o uso de bases de dados censitárias, como o Censo Educacional, coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). O presente estudo também apresenta um estudo de caso que demonstra uma das aplicações possíveis dessa arquitetura e evidencia o valor potencial no uso dela para a gestão educacional no país. Concluímos assim, que a arquitetura proposta é um instrumento útil para suportar plataformas de avaliação e acompanhamento de políticas públicas educacionais, contribuindo para a construção de um Governo Inteligente e para promover melhorias na qualidade da educação brasileira.*

## **1. Introdução**

O Plano Nacional de Educação (PNE) brasileiro pretende orientar o desenvolvimento da educação no país por 10 anos [Macena et al. 2018]. O PNE estabelece metas, estratégias e ações para a melhoria da qualidade da educação em todas as etapas e modalidades de ensino, da educação infantil ao ensino superior, bem como para a formação de professores e para a gestão democrática das escolas. O monitoramento do PNE é uma atividade fundamental para manter a gestão na direção de cumprir as metas estabelecidas [Vinente and Duarte 2015] e para possibilitar ajustes ou mudanças de trajetória nas estratégias de gestão com o intuito de alcançar o plano definido.

Visando possibilitar o monitoramento do PNE e auxiliar na tomada de decisões baseadas em evidências, é necessário disponibilizar soluções capazes de fornecer um conjunto de dados confiável e capaz de gerar tais evidências. Assim, a disponibilização de uma arquitetura de dados para acompanhamento dos Indicadores do PNE em formato acessível, tanto para gestores educacionais públicos quanto para privados, pode ser uma ferramenta valiosa para nutrir esses agentes de informações para avaliar os resultados das políticas educacionais adotadas no país. Estudos anteriores já demonstraram a eficácia de técnicas de análise de dados visando otimizar o sistema educacional municipal, como o realizado por [Júnior et al. 2019] para a cidade de Maceió-AL. No entanto, ainda são necessários esforços para disponibilizar soluções que permitam esse tipo de análise para todos os municípios do Brasil.

Considerando a necessidade de lidar com diferentes fontes de dados em condições diversas de formato, estrutura e nível de qualidade, torna-se necessário discutir o processo de geração dos indicadores do PNE, que envolve a jornada para obtenção e transformação dos dados brutos oriundos dos sistemas governamentais. Para que os indicadores produzidos possam capacitar tomadas de decisões confiáveis ao nível municipal, é fundamental evitar o uso de fontes de dados geradas a partir de consultas amostrais. Em vez disso, devem ser utilizadas bases de dados censitárias, como o Censo Educacional [Diniz 1999] coordenado pelo Instituto Nacional de Estudos e Pesquisas Educacionais (INEP), órgão autônomo regulamentado por instrumentos normativos que instituem a obrigatoriedade, os prazos, os responsáveis e suas responsabilidades, bem como os procedimentos para realização de todo o processo de coleta de dados.<sup>1</sup>

---

<sup>1</sup><https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-escolar>

Nesse contexto, o presente trabalho apresenta uma arquitetura de dados educacionais que utiliza dados abertos para apoiar a gestão educacional baseada em evidências. Visando a construção de um Governo Inteligente, essa arquitetura disponibiliza os dados dos indicadores instituídos pelo PNE às secretarias de educação e à população, possibilitando o cruzamento de informações de diversas fontes e variáveis que podem estar influenciando no alcance das metas estabelecidas. Dessa forma, a arquitetura apresenta-se como um instrumento útil no processo de tomada de decisão, possibilitando o desenvolvimento de sistemas de apoio à gestão educacional, além de aumentar a transparência e divulgação do impacto causado pelos investimentos na área.

Este documento encontra-se organizado da seguinte maneira: a Seção 2 apresenta alguns trabalhos relacionados ao uso de dados abertos, uso de dados para tomada de decisão e *pipeline* de processamento de dados educacionais e populacionais; na Seção 3 é apresentada arquitetura proposta, suas fonte de dados, ferramentas utilizadas e o processo de aquisição e tratamento de dados seguidos; na Seção 4.1 são demonstrados alguns dos resultados que já podem ser atingidos com o uso da arquitetura em um estudo de caso onde realizamos o cruzamento dos dados do PNE com outros indicadores selecionados a partir da base de perfil municipal MUNIC (Pesquisa de Informações Básicas Municipais) divulgados pelo Instituto Brasileiro de Geografia e Estatística (IBGE); e, por fim, a Seção 5 traz as considerações finais e trabalhos futuros.

## 2. Trabalhos Relacionados

Alguns trabalhos já foram produzidos sobre as bases educacionais oferecidas pelo INEP. Em [Conte 2019], foram utilizadas bases de perfil socioeconômico e resultados do Exame Nacional do Ensino Médio (ENEM), permitindo, assim, a realização de procedimentos de limpeza e organização dos dados para fornecer conhecimento sobre a base do exame na edição do ano de 2015. Como destacado pelo autor, é extremamente importante associar esse conhecimento adquirido com outras bases fornecidas, como a Prova Brasil, o Exame Nacional de Desempenho de Estudantes (ENADE) e o Sistema de Avaliação da Educação Básica (SAEB).

Em [Wanderley 2021], foi apresentado alguns indicadores para monitorar instituições escolares do estado do Acre. As fontes de dados utilizadas para a criação e apontamento desses indicadores foram as bases de dados educacionais fornecidas pelo INEP. Após o processamento desses dados, os resultados dos indicadores foram utilizados para a criação de painéis utilizando a ferramenta *Power BI*.

A importância dos *data lakes* lógicos como plataforma para dados governamentais em sociedades e cidades inteligentes é discutido no artigo [Ramos et al. 2022]. O autor destaca a necessidade de gerenciar grandes volumes de dados provenientes de várias fontes para promover a tomada de decisão informada e a melhoria dos serviços governamentais. O artigo também descreve as principais características de um *data lake* lógico, incluindo a capacidade de lidar com dados estruturados e não estruturados, escalabilidade e flexibilidade.

Já em [Rodrigues and Maciel 2022], destaca-se a importância dos dados abertos educacionais para a tomada de decisões informadas e a melhoria da qualidade da educação. O método proposto envolve a extração de dados brutos de várias fontes, a transformação dos dados para um formato padronizado e a carga dos dados em um re-

positório centralizado. O artigo também apresenta uma arquitetura de software para implementar o processo *ETL* (Extração, Transformação e Carga) e discute a importância da segurança dos dados e da privacidade dos usuários. Tendo como objetivo do método proposto é melhorar a eficiência da captura e compartilhamento de dados educacionais abertos, permitindo que os usuários acessem dados atualizados e relevantes para melhorar a educação.

O artigo apresentado em [Albuquerque et al. 2022] descreve um conjunto de dados que fornece estimativas populacionais desagregadas por município e idade no Brasil para o período de 2014 a 2020. Para criar esse conjunto de dados, os autores utilizaram uma metodologia que combina informações do Censo Demográfico de 2010 com as projeções populacionais do IBGE. O artigo detalha a metodologia utilizada, incluindo os ajustes realizados nas projeções populacionais para garantir a consistência temporal e espacial dos dados.

Uma abordagem para a junção histórica dos dados do censo da educação básica de 2011 a 2021 em seu novo formato da versão jan/2022, é apresentado em [Barros et al. 2022]. Seu objetivo foi o de obter uma base de dados resultante com pouco descarte de colunas que foram sendo descontinuados do Censo Educacional ao longo dos anos. O trabalho detalha o *pipeline* executado e as parametrizações realizadas para alcançar esse resultado. Além disso, destaca-se que essa abordagem pode ser replicada para determinados anos alvo de projetos educacionais.

Os estudos mencionados foram úteis para iniciar o processo de concepção e elaboração de uma arquitetura que possa lidar com grandes quantidades de dados públicos e que permita o monitoramento unificado das metas e indicadores do PNE de todos os municípios. Essa arquitetura deve fornecer uma visão geral da situação educacional ao nível federal e ser transparente desde a coleta até o fornecimento dos dados. Assim, este trabalho busca preencher a lacuna de uma arquitetura que forneça os serviços e dados necessários para esse monitoramento.

### **3. Arquitetura de Dados Educacionais como Plataforma para Governo Inteligente**

Este estudo propõe uma arquitetura de dados para o acompanhamento dos indicadores do PNE ao longo do ciclo de 2014 a 2024. Entre as etapas para a construção dessa arquitetura e serão abordadas neste artigo, destacam-se a aquisição, armazenamento, processamento e disponibilização dos indicadores e variáveis de interesse.

A Figura 1 apresenta a arquitetura proposta, onde do lado esquerdo pode-se observar as fontes de dados públicas atualmente utilizadas, que serão detalhadas na Seção 3.1. Ao lado direito, temos o fluxo conceitual de como os dados são implantados e processados.

Após o processamento, os dados são armazenados em um banco de dados relacional no PostgreSQL<sup>2</sup>. Esse banco é dividido em dois esquemas: *staging* e *serving*.

No esquema *staging*, são armazenados os dados brutos provenientes de outras fontes de dados, como arquivos Comma-separated values (CSV) originado do IBGE e do

---

<sup>2</sup><https://www.postgresql.org/>

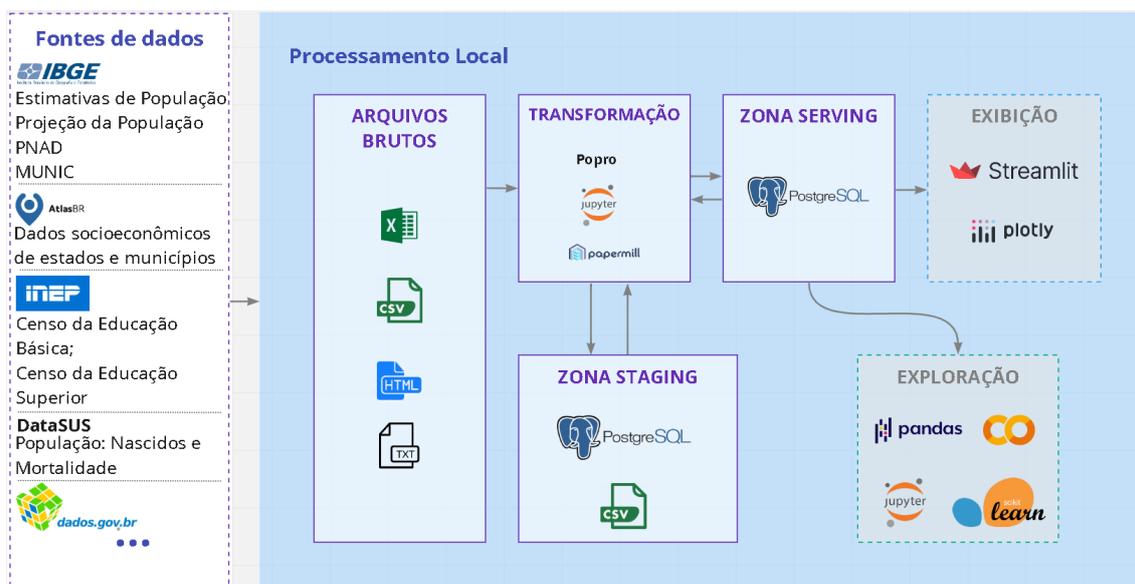


Figura 1. Estrutura da arquitetura de dados educacionais proposta

INEP, e dados de *Application Programming Interface* (API) do Ipeadata<sup>3</sup>.

No esquema *servicing*, armazenam-se as tabelas com dados que foram processados e filtrados conforme os interesses de análise, estando aptas para consumo. As tabelas nessa zona são construídas com dados num formato que capacite análises e filtros por ano e níveis geográficos como estado e município. Estas tabelas também fornecem informações específicas, como os valores já calculados dos indicadores das metas do PNE, as variáveis necessárias para os cálculos e outras informações úteis, como dados socio-econômicos, que capacita uma análise multifacetada. Deste modo, esta zona fornece dados que estão prontos para serem consumidos por ferramentas de visualização de dados e Business Intelligence (BI), que estão presentes nas zonas de EXPLORAÇÃO e EXIBIÇÃO da arquitetura. Um exemplo de como esses dados podem ser consumidos será apresentado no estudo de caso da Seção 4.1.

Vale destacar que as ferramentas tecnológicas utilizadas na criação e instanciação da arquitetura são gratuitas e de código aberto, o que permite a replicação da arquitetura sem custos adicionais. Além do fator de código aberto, tais ferramentas possuiriam uma melhor curva de aprendizado e apoio a implementação de ambiente *on premise*, por caso deseje ser replicado é possível sua aplicação em ambientes como *Amazon Web service* ou *Google Cloud plataforma*.

### 3.1. Aquisição e Processamento de dados

Os dados utilizados atualmente vêm em sua totalidade de fontes Governamentais como o INEP<sup>4</sup>, IBGE<sup>5</sup>, Instituto de Pesquisa Econômica Aplicada (Ipea)<sup>6</sup>, Departamento de Informática do Sistema Único de Saúde (DATASUS)<sup>7</sup> e de organizações como o Atlas Bra-

<sup>3</sup><http://www.ipeadata.gov.br/api/>

<sup>4</sup><https://www.gov.br/inep/pt-br>

<sup>5</sup><https://www.ibge.gov.br>

<sup>6</sup><https://www.ipea.gov.br/portal/>

<sup>7</sup><https://datasus.saude.gov.br>

sil<sup>8</sup> mantida pelos órgãos Fundação João Pinheiro (FJP), Ipea e o Programa das Nações Unidas para o Desenvolvimento (PNUD).

Durante a etapa de aquisição de dados, foi possível notar que a disponibilização desses dados sofreu alterações desde o início do projeto em novembro de 2021. Primeiramente, algumas colunas foram atualizadas devido ao tipo de dados e à nomenclatura das colunas informadas. Em janeiro de 2022, foi divulgada a intenção de disponibilizar uma nova versão desses dados para se adequar à Lei Geral de Proteção de Dados (LGPD)<sup>9</sup>. Os dados que antes eram disponibilizados de forma desagregada e individual, com informações anonimizadas de alunos e professores, agora são apresentados em formato sintético, reunidos em um único arquivo e com registros a nível de instituições. Para o ensino superior, é fornecida uma tabela adicional com dados já agregados sobre os cursos. No entanto, essa agregação impede análises mais específicas.

A publicação desse novo formato de dados acarretou prejuízos, como a impossibilidade do cálculo ou atualização de alguns indicadores, pois a versão anterior foi descontinuada em 2020 e alguns indicadores dependem de filtros compostos de variáveis, como idade do aluno e etapa escolar específica. Algumas dessas variáveis necessárias não estão disponíveis no atual modelo de dados publicado. A justificativa oficial divulgada pelo INEP para a mudança estrutural da base de dados foi a necessidade de mitigar o risco de rastreamento de alunos ou professores com base nas informações relatadas no censo em seu formato anterior. Mais detalhes sobre o estudo realizado para a adequação podem ser encontrados no portal de microdados do INEP<sup>10</sup>. No entanto, dados gerais, como tipo de etapa de ensino ou qual nível de qualificação realizada pelo docente, são categorias necessárias para o cálculo de indicadores. Diante disso, para esses indicadores, a plataforma apresentará os resultados até o ano de 2020. Para as bases educacionais, foi aplicado o processo descrito em [Barros et al. 2022] para o processamento das bases vindas do INEP, sendo elas o Censo Educacional da Educação Básica e o Censo da Educação Superior.

No que diz respeito aos dados socioeconômicos e populacionais, realizou-se a coleta nos sites oficiais, como o IBGE e DATASUS, que fornecem os insumos essenciais para a criação das tabelas. Os sites do Ipea e PNUD servem como fontes para o enriquecimento dos dados.

As fontes de dados governamentais disponibilizadas pelo IBGE não fornecem estimativas populacionais segmentadas por idade para municípios, o que é fundamental para o cálculo de diversos indicadores do PNE. A divulgação de estimativas populacionais segmentadas por idade para anos posteriores ao último Censo Demográfico realizada pelo IBGE é restrita ao nível de Estado Federativo, o que representa um desafio para criação de indicadores demográficos ao nível municipal.

Para suprir essa lacuna, utilizou-se o processo descrito em [Albuquerque et al. 2022] para gerar as estimativas populacionais segmentadas tanto por idade quanto por municípios, sendo essa etapa imprescindível para o contexto do monitoramento de indicadores e metas do PNE.

Os dados socioeconômicos vindos do PNUD e Ipea apresentam granularidade

---

<sup>8</sup><http://www.atlasbrasil.org.br/>

<sup>9</sup><https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados>

<sup>10</sup><https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados>

anual e municipal, possibilitando o cruzamento com os dados educacionais para viabilizar o cálculo dos indicadores que envolvem dados populacionais.

### **3.2. Dicionário de dados**

Para cada tabela produzida, foi criado um novo dicionário de dados para auxiliar na utilização e montagem de novas tabelas. Sua produção foi baseada nos dicionários originais fornecidos pelos órgãos oficiais e é acrescido dos novos campos calculados ou modificações aplicadas pelos autores. O dicionário completo pode ser acessado em <https://aiboxlab-pne.github.io/dados/dict/serving/> e <https://aiboxlab-pne.github.io/dados/dict/staging/>.

## **4. Resultados**

A principal contribuição deste trabalho é a disponibilização da arquitetura de dados para consumo por sistemas e ferramentas interessados em serviços de dados para o monitoramento do PNE ou outras atividades correlatas ligadas à melhoria da qualidade da educação no Brasil.

Com o intuito de demonstrar a aplicação da arquitetura de dados desenvolvida, realizamos um experimento que consistiu no cruzamento dos dados do PNE com outras variáveis educacionais apuradas na esfera municipal.

Cada meta do PNE possui um ou mais indicadores associados a ela, que aferem quantitativamente desde o percentual de atendimento de uma faixa etária na escola até a eficiência na gestão escolar. Uma vez que os indicadores educacionais são impactados por fatores diversos [De Witte et al. 2013], convém analisar a relação e possível influência de diferentes variáveis sobre os indicadores do PNE.

Esse tipo de análise pode contribuir para aprimorar as políticas públicas, já que permite identificar fatores que podem estar influenciando no alcance das metas estabelecidas pelo PNE em âmbito federal.

### **4.1. Estudo de Caso**

Para demonstrar a aplicação prática da arquitetura de dados desenvolvida, foi realizado um estudo de caso que será apresentado nesta Seção, visando evidenciar as possibilidades de extração de utilidade a partir dessa arquitetura, auxiliando as equipes de gestão educacional a tomarem decisões baseadas em evidências.

Nesse estudo de caso, utilizamos a base de dados de perfil municipal (MUNIC)<sup>11</sup>, tendo como referência o Indicador 1A do Plano Nacional de Educação (PNE). A análise comparada dos múltiplos indicadores municipais permitiu avaliar o impacto das políticas educacionais implementadas pelos municípios. Ademais, a base MUNIC apresenta-se como uma fonte valiosa para a análise cruzada dos fatores que podem estar influenciando no alcance das metas estabelecidas em âmbito federal pelo PNE.

Para identificar alguns fatores de influência no Indicador 1A do PNE (Percentual da população de 4 a 5 anos que frequenta a escola/creche), realizou-se um experimento a partir dos indicadores educacionais da base de dados MUNIC. O histórico de dados

---

<sup>11</sup><https://cidades.ibge.gov.br/brasil/pesquisa/1/74454>

educacionais nessa base contemplou os anos de 2014, 2018 e 2021. Para o ano de 2021, foram disponibilizadas 102 variáveis booleanas, embora nem todas essas variáveis estivessem presentes nas bases de dados dos anos 2018 e 2014. De forma geral, essas variáveis indicam se há a incidência ou ocorrência de um dado fator.

Uma resposta verdadeira ao fator é sempre algo semanticamente positivo. Nesse sentido, quanto mais respostas marcadas como verdadeiras, melhor é a situação geral do município. Para possibilitar a avaliação de um município a partir de um conjunto dessas variáveis, foi utilizado a razão entre as respostas verdadeiras e o total de indicadores, resultando no Fator Educacional MUNIC (FEM), o que representa um desafio para criação de indicadores demográficos ao nível municipal.

Partindo dessa base, foram filtradas variáveis com equivalência histórica que possuíam alguma relação tanto com o Indicador 1A, quanto as 13 variáveis pré-selecionadas. Em seguida, ao analisar o comportamento do FEM em relação ao Indicador 1A do PNE, foi identificada uma baixa influência entre as duas métricas, o que demandou uma seleção mais criteriosa a partir das 13 variáveis para formar um subconjunto que explicasse melhor o comportamento do Indicador 1A do PNE. Para realizar essa seleção, foi aplicado um processo para busca de subconjuntos com variáveis, mapeando todas as combinações possíveis de variáveis para poder ser gerado um subconjunto mínimo de 5 elementos, esse processo gerou um total de 7.099 subconjuntos a serem analisados.

Todos os subconjuntos foram analisados considerando sua correlação com o indicador 1A do PNE e seu comportamento histórico. Para explicar o processo de escolha do subconjunto, chamamos de comportamento em “sintonia” quando um município apresenta um comportamento entre 2014 e 2021 de subida, descida ou inalteração do *FEM* acompanhado do mesmo comportamento no Indicador 1A do PNE.

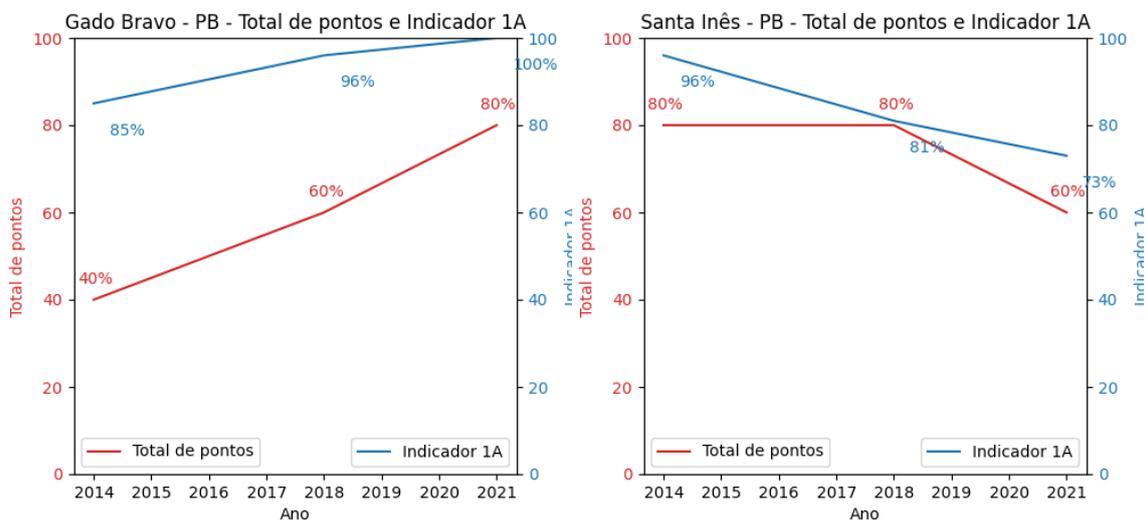
De todas as 7.099 combinações possíveis, o subconjunto de variáveis que gerou maior volume de municípios com comportamento de subida e descida em sintonia entre o *FEM* e o Indicador 1A do PNE, foi composto pelas 5 variáveis listadas abaixo.

1. Plano Municipal de Educação - existência
2. Plano de Carreira para o Magistério - existência
3. Conselho de Educação - existência
4. Conselho de alimentação escolar - existência
5. Conselho de transporte escolar - existência

Considerando o subconjunto escolhido em uma análise dos 233 municípios da Paraíba, foi observado que 96 municípios apresentaram comportamento de subida em sintonia, 2 municípios tiveram comportamento de queda, e 1 município permaneceu estacionário. Assim 109 municípios identificados como em sintonia (44,3% do total) entre o *FEM* e o indicador 1A do PNE. Por outro lado, os demais 121 municípios apresentaram comportamento diferente entre o *FEM* e o Indicador 1A do PNE.

Na Figura 2, é ilustrada uma amostra da relação entre o Indicador 1A do PNE e o fator *FEM* a partir de 2014, demonstrando um movimento de sintonia na melhora ou piora deles, para os municípios de Gado Bravo e Santa Inês, respectivamente. A partir da considerável relação da composição do subconjunto de indicadores municipais e o Indicador 1A do PNE, os gestores municipais têm um ponto de partida alinhado aos

interesses do Plano Nacional de Educação, sobre quais variáveis requerem maior atenção dentre as 5 existentes no subconjunto *FEM*.



**Figura 2. Amostra da relação entre o Indicador 1A do PNE e o Fator Educacional MUNIC (FEM) a partir de 2014.**

Após o estudo, é importante ressaltar que uma arquitetura de dados bem estruturada e integrada pode, portanto, potencializar análises e, conseqüentemente, trazer mais transparência e efetividade para a definição e aplicação das políticas públicas educacionais.

Ressaltando que este é apenas um estudo de caso no qual é possível cruzar algumas das bases disponíveis na plataforma. Os dados disponíveis representam todo o território brasileiro.

## 5. Conclusão

O presente artigo apresentou o desenvolvimento de uma arquitetura de dados educacionais usando dados abertos governamentais, solucionando a observada necessidade de dados para sustentar serviços e plataformas de apoio a uma gestão educacional baseada em evidências. Observa-se assim a capacidade de uso dessa arquitetura para acompanhar o andamento das metas, estratégias e ações do PNE e, conseqüentemente, apoiar o processo de tomada de decisão.

A partir do apoio oferecido por essa arquitetura de dados, as secretarias de educação e a população em geral têm acesso aos dados dos indicadores do PNE e demais dados por metodologias censitárias (não amostrais) e todas as variáveis necessárias para a formação destes indicadores. Isso pode auxiliar na avaliação e acompanhamento das políticas educacionais adotadas, bem como contribuir na construção de um Governo Inteligente. O estudo de caso apresentado na Seção 4.1 demonstrou uma das aplicações possíveis dessa arquitetura. Diante disso, evidencia-se o valor potencial no uso dessa arquitetura por gestores educacionais para promover melhorias na qualidade da educação brasileira.

## 5.1. Trabalhos Futuros

A arquitetura de dados construída neste estudo pode ser expandida para incluir novas metas e serviços, além daqueles já contemplados e apresentados ao longo deste artigo, sendo um esforço a ser dedicado para trabalhos futuros. Além disso, é possível incorporar outras bases de dados que tenham relação direta ou indireta com os indicadores do PNE, como dados de empregabilidade ou de violência. Esses dados podem propiciar grupos de estudos e trabalhos conjuntos entre a secretaria de educação e outras secretarias, em sintonia com as metas definidas pelo PNE. Por fim, é possível realizar melhorias técnicas, como a automação do processo de transformação do pipeline de dados, tornando-o mais robusto e escalável.

## Referências

- Albuquerque, A., Barros, A., Alencar, A., Nascimento, A., Bittencourt, I., and Mello, R. (2022). Dataset de estimativas populacionais desagregada por município e idade 2014-2020. In *Anais do IV Dataset Showcase Workshop*, pages 25–34, Porto Alegre, RS, Brasil. SBC.
- Barros, A. N., Alencar, A., Nascimento, A., de Albuquerque, A. F., and Mello, R. F. (2022). Elaboração do conjunto de dados agregados do censo da educação básica. In *Anais do IV Dataset Showcase Workshop*, pages 35–45. SBC.
- Conte, V. d. S. (2019). Mineração de dados educacionais para avaliar os fatores que influenciam no desempenho de candidatos do enem.
- De Witte, K., Cabus, S., Thyssen, G., Groot, W., and van den Brink, H. M. (2013). A critical review of the literature on school dropout. *Educational Research Review*, 10:13–28.
- Diniz, E. (1999). O censo escolar. *Revista Brasileira de Estudos Pedagógicos*, 80(194).
- Júnior, O. d. G. F., Rodrigues, W. R. M., Barbirato, J. C. C., and de Barros Costa, E. (2019). Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de maceió. *RENOTE*, 17(1):296–305.
- Macena, J. d. O., Justino, L. R. P., and Capellini, V. L. M. F. (2018). O plano nacional de educação 2014–2024 e os desafios para a educação especial na perspectiva de uma cultura inclusiva. *Ensaio: Avaliação e Políticas Públicas em Educação*, 26:1283–1302.
- Ramos, G. S., Fernandes, D., Coelho, J. A. P. d. M., and Aquino, A. L. (2022). Data lakes lógicos como plataformas para dados governamentais em sociedades e cidades inteligentes. In *Anais do X Workshop de Computação Aplicada em Governo Eletrônico*, pages 215–226. SBC.
- Rodrigues, F. A. and Maciel, C. (2022). Um método para captura e compartilhamento de dados abertos educacionais via um processo etl. In *Anais do X Workshop de Computação Aplicada em Governo Eletrônico*, pages 133–144. SBC.
- Vinente, S. and Duarte, M. (2015). O plano nacional de educação (2014-2024) e a garantia de um sistema educacional inclusivo: possibilidade ou utopia. *Olhares: Revista do Departamento de Educação da Unifesp*, 3(2):133–151.

Wanderley, P. F. (2021). Uso de business intelligence para avaliação de indicadores de desempenho na educação básica: um estudo de caso no estado do acre.