

1. Resumo executivo

Contextualização

A confiança nas instituições eleitorais está ameaçada por desinformação e narrativas de fraude. A transparência e o acesso a dados eleitorais confiáveis são cruciais para reconstruir a confiança pública e fortalecer a democracia. O Tribunal Superior Eleitoral (TSE) disponibiliza microdados desde 2009 através do Repositório de Dados Eleitorais (RDE), mas o acesso direto exige conhecimento técnico.

Objetivos do Projeto

O Portal da Classe Política é uma plataforma digital interativa que transforma dados brutos do TSE em informações acessíveis, organizando-os em um banco relacional com visualizações, filtros e indicadores. Os principais objetivos são: Democratizar o acesso aos dados eleitorais para diversos públicos, fomentar a produção científica em Ciência Política com dados organizados, promover a educação política através da visualização de informações e apoiar a confiança nas instituições democráticas via transparência de dados.

Metodologia

A metodologia adotada no Portal da Classe Política combina técnicas avançadas de engenharia de dados com fundamentos da Ciência Política empírica. O projeto parte de uma base robusta que consolida todas as eleições brasileiras entre 1998 e 2022, totalizando mais de 2,8 milhões de candidaturas e cerca de 1,8 milhão de indivíduos únicos. A partir dos dados brutos disponibilizados pelo TSE, foi implementado um processo rigoroso de padronização e tratamento, que envolve a estruturação de um banco de dados relacional com arquitetura entidade-relacionamento normalizada, escalável e preparada para cruzamentos complexos. Com a base limpa e unificada, foram desenvolvidos indicadores analíticos que automatizam o cálculo de mais de 16 índices clássicos da literatura de Ciência Política e Sociologia Eleitoral, além de métricas autorais elaboradas pela equipe técnica.

Impactos gerados

O Portal da Classe Política gera impactos em frentes estratégicas para o fortalecimento democrático. No campo da transparência eleitoral, amplia o acesso público a dados organizados, auditáveis e apresentados de forma clara, reduzindo as barreiras técnicas que limitavam o uso dos microdados do TSE. No âmbito do engajamento cívico, disponibiliza ferramentas intuitivas que permitem aos cidadãos acompanhar o histórico completo de candidaturas, votações e patrimônio declarado dos políticos brasileiros. Na Ciência Política, a base de dados relacional padronizada possibilita análises comparadas, replicáveis e historicamente estruturadas, fomentando uma agenda de pesquisa empírica robusta e alinhada à ciência aberta. Por fim, o projeto também contribui para a educação política ao oferecer visualizações interativas que auxiliam estudantes, professores e interessados a compreenderem melhor o funcionamento das

eleições, o comportamento das elites políticas e os padrões do sistema eleitoral brasileiro.

Portal da Classe Política: processamento, visualização e transparência de *bigdata* eleitoral no Brasil

Nilton Sainz

Adriano Codato

Bruno Bolognesi

2. Introdução

Nas democracias contemporâneas, a confiança pública nas instituições eleitorais tornou-se um aspecto sensível e em constante disputa. Diversos países vêm enfrentando a erosão dessa confiança, frequentemente alimentada por desinformação e por narrativas infundadas de fraude. No Brasil, esse cenário é ilustrado por pesquisas de opinião recentes: em 2024, apenas 40% dos brasileiros declararam confiar muito nas urnas eletrônicas (Castro, 2022), refletindo um clima de ceticismo não apenas em relação ao sistema de votação, mas também à Justiça Eleitoral como um todo. Paralelamente, cerca de 63% da população afirma que as notícias falsas mais comuns dizem respeito a campanhas políticas, o que dificulta a compreensão dos fatos e alimenta suspeitas infundadas (Quase 90% dos brasileiros admitem ter acreditado em *fake news*, 2024). Esse ambiente de desinformação e dúvida sistemática contribui para o questionamento da integridade dos pleitos e da honestidade dos atores políticos, enfraquecendo a legitimidade do regime democrático.

Diante desse quadro, a transparência e o acesso público aos dados eleitorais emergem como instrumentos centrais para a reconstrução da confiança cívica. A abertura de informações possibilita que cidadãos, jornalistas e pesquisadores

acompanhem de forma independente o funcionamento do processo eleitoral, o que limita a propagação de boatos e reduz as percepções de fraude. Como argumenta Kaplan (2006), a transparência é fundamental porque amplia a credibilidade do sistema de votação ao mesmo tempo em que inibe práticas irregulares. Quando as instituições disponibilizam dados completos e auditáveis — como resultados por urna, informações detalhadas sobre candidaturas ou receitas de campanha —, criam-se oportunidades para controle social e fiscalização direta. Pesquisas comparadas apontam que a ausência de abertura e clareza nos procedimentos eleitorais tende a minar a confiança da população (Duenas-Cid et al., 2024), enquanto iniciativas que permitem auditoria pública dos resultados têm potencial para fortalecê-la.

No Brasil, o direito de acesso à informação é garantido constitucionalmente e regulado pela Lei de Acesso à Informação (Lei nº 12.527/2011), que estabelece a publicidade como regra e o sigilo como exceção. A LAI determina, entre outros pontos, a divulgação proativa de dados de interesse público, independentemente de solicitação formal, bem como o uso de meios tecnológicos para assegurar o acesso. Os artigos 4º e 5º da norma reforçam a importância de fomentar uma cultura de transparência e estimular o controle social sobre a administração pública. Isso implica um dever institucional não apenas de responder a demandas, mas de disponibilizar ativamente dados em formatos acessíveis e compreensíveis.

Foi nesse contexto normativo e político que a Justiça Eleitoral brasileira passou a adotar uma postura mais ativa na transparência de seus processos. Um marco importante foi a criação do Repositório de Dados Eleitorais (RDE), lançado em 2009 como uma parceria entre a Secretaria de Tecnologia da Informação e a equipe de estatística do Tribunal Superior Eleitoral (TSE). Desde então, o repositório vem sendo continuamente ampliado e atualizado, consolidando-se como uma referência nacional em transparência eleitoral (Bezerra Da Silva et al., 2020).

O RDE reúne um volume expressivo de dados públicos que cobrem a totalidade dos pleitos realizados no Brasil desde 1945, contemplando diferentes períodos políticos — da redemocratização às eleições contemporâneas. Os usuários podem consultar planilhas de dados de resultados eleitorais por cargo e localidade, verificar quantos votos um candidato recebeu, acessar seu perfil sociodemográfico (gênero,

idade, escolaridade, profissão, município de nascimento), bem como obter informações sobre coligações partidárias e situação das candidaturas. A partir de 2014, a declaração racial dos candidatos também passou a integrar os conjuntos de dados, ampliando as possibilidades de análise sobre representatividade e desigualdade eleitoral.

Além dos dados sobre candidatos e votações, o repositório inclui registros sobre o eleitorado, com estatísticas por sexo, faixa etária, grau de instrução, uso de nome social e presença de eleitores com deficiência, detalhados por zona eleitoral. Também estão disponíveis informações sobre partidos políticos, pesquisas de intenção de voto, prestações de contas de campanhas (desde 2002), contabilidade anual dos partidos, e registros sobre processos judiciais eleitorais, como ações, decisões e recursos. Um dos conjuntos mais robustos diz respeito aos boletins de urna, disponíveis de forma detalhada desde 1998, que permitem auditorias independentes ao associar cada resultado à seção e à urna correspondente, possibilitando a verificação da integridade do processo de totalização.

A justificativa para a existência de um repositório tão amplo está ancorada nos princípios de governo aberto e no arcabouço legal de transparência que rege a administração pública brasileira (Brandt et al., 2018). O TSE, ciente de seu papel institucional na garantia da integridade do processo democrático, adotou uma postura proativa ao disponibilizar dados que permitem o escrutínio público de todas as etapas da eleição — do registro de candidaturas à contagem final dos votos. Com isso, o repositório não apenas cumpre exigências legais de transparência ativa, mas também fortalece a legitimidade do sistema eleitoral ao convidar a sociedade a verificar, com base em evidências concretas, a lisura de cada pleito.

O impacto prático desse repositório é particularmente significativo para a Ciência Política, que encontra nele uma base estruturante para investigações empíricas sobre o sistema político brasileiro. Pesquisadores que se dedicam aos estudos eleitorais utilizam os microdados de votação, candidaturas e quocientes para analisar padrões de competição partidária, volatilidade e fragmentação do sistema. Aqueles voltados à representação política mobilizam dados de gênero, raça e ocupação para investigar desigualdades no acesso às arenas políticas e a eficácia de políticas de

inclusão. Já os estudiosos de elites políticas exploram informações sobre patrimônio, reeleições e mudanças de partido para compreender a trajetória e permanência de determinados grupos no poder. Na área de comportamento político, os dados de voto por seção, cruzados com características do eleitorado, permitem análises contextuais e a construção de modelos explicativos sobre decisão de voto, preferências e percepção dos candidatos.

A relevância do repositório, no entanto, transcende os limites da academia. Jornalistas investigativos, organizações da sociedade civil e profissionais do campo político utilizam as bases para rastrear vínculos entre candidatos e doadores, verificar a coerência programática de campanhas ou monitorar o desempenho de candidaturas sub-representadas. Cidadãos engajados, estudantes, educadores encontram no repositório do TSE um instrumento que favorece a compreensão crítica do sistema eleitoral. Sem esse esforço de centralização, grande parte dos dados estaria fragmentada em arquivos administrativos, publicações pouco restritas que dependeria de procedimentos burocráticos para ser obtida.

3. Portal da Classe Política: transparência, acessibilidade e confiança nos dados eleitorais

Com o objetivo de aprofundar a transparência e democratizar o acesso à informação eleitoral no Brasil, surgiu o projeto “Portal da Classe Política”. Trata-se de uma iniciativa técnico-científica voltada à unificação, padronização e disponibilização pública dos dados eleitorais produzidos pelo Tribunal Superior Eleitoral (TSE), por meio de uma plataforma online acessível, interativa e intuitiva. O portal funciona como uma ponte entre o vasto repositório de dados brutos do TSE e os diferentes públicos interessados — como pesquisadores, jornalistas, estudantes e cidadãos — traduzindo essa massa de informações em visualizações compreensíveis, indicadores analíticos e recursos de exploração personalizados.

Embora o Brasil disponha de dados eleitorais desde 1945, foi a partir de 1998 que os bancos passaram a apresentar maior consistência, padronização e integração

digital. Por isso, o Portal foca inicialmente na consolidação das 14 eleições gerais e municipais realizadas entre 1998 e 2024, organizando seus dados em um banco relacional único. Esse processo lida com desafios próprios do *Big Data*, ao integrar milhões de registros provenientes de bases heterogêneas, aplicando técnicas de limpeza, normalização e validação para assegurar interoperabilidade. O resultado é uma infraestrutura de dados que permite consultas simples e análises sofisticadas sem exigir conhecimento técnico aprofundado dos usuários.

A interface do Portal da Classe Política permite, por exemplo, consultar o histórico eleitoral de um político específico, visualizar mapas dinâmicos de votação por localidade, acompanhar a evolução de desempenho partidário ao longo do tempo ou analisar mudanças no perfil social e econômico da elite política brasileira. Para isso, o projeto não se limita a apresentar dados brutos, mas desenvolve indicadores sintéticos como o Número Efetivo de Partidos (NEP), a taxa de reeleição, renovação líquida, o quociente Partidário e eleitoral, volatilidade eleitoral de Pedersen, índice de paridade eleitoral de gênero, entre outros, permitindo uma leitura qualificada e comparativa das dinâmicas eleitorais.

A proposta do portal está alinhada aos princípios de Governo Aberto e à agenda de Dados Abertos Governamentais, conforme discutido na literatura especializada. Autores como Janssen et al. (2012) e Ruijter et al. (2020) destacam que a transparência digital efetiva exige não apenas a disponibilização dos dados, mas também sua inteligibilidade e aplicabilidade social. Ao transformar dados complexos em narrativas compreensíveis, o portal incorpora os princípios discutidos por Dourish e Cruz (2023), ao tratar os dados como construções que precisam ser contextualizadas, interpretadas e narradas para se tornarem significativas. Nesse sentido, o projeto também contribui para valorizar e projetar a Ciência Política brasileira, ao evidenciar, de forma aplicada, como o conhecimento acumulado pela área pode ser mobilizado para interpretar fenômenos políticos complexos, fortalecer a transparência pública e aproximar a sociedade dos dados e instituições que estruturam o regime democrático.

Ademais, o portal busca fomentar a *accountability* democrática. Ao tornar o histórico de candidaturas, as prestações de contas e os resultados de votação acessíveis a qualquer pessoa, o projeto permite que eleitores avaliem a coerência das

trajetórias políticas, identifiquem padrões de desempenho eleitoral, acompanhem os estágios de desigualdade política entre grupos minoritários (como gênero e raça) e fiscalizem declaração de bens e recursos de campanha. Dessa forma, contribui para um voto mais consciente e informado, ao mesmo tempo em que amplia a visibilidade das instituições e reforça a confiança nas regras do jogo democrático.

Em um contexto de crescente polarização e circulação de narrativas infundadas sobre fraudes eleitorais, o Portal da Classe Política também se apresenta como um antídoto contra a desinformação e a opacidade, ao permitir a visualização estruturada dos dados consolidados pela Justiça Eleitoral e ao se somar como mais uma camada independente de auditoria cidadã dos dados públicos. A inteligibilidade da informação e a possibilidade de auditoria pública são condições fundamentais para preservar a estabilidade do sistema político em tempos onde a desconfiança alcança todos os estágios do ciclo eleitoral, como argumentou Duenas-Cid *et al.* (2024). Nesse sentido, o Portal caminha para ser inteiramente auditável, oferecendo à comunidade científica e à sociedade acesso via API (em SQL) e pacote em R, acompanhado de documentação técnica, *codebook* e arquitetura de dados aberta, alinhando-se aos princípios da Ciência Aberta e da transparência metodológica.

O projeto está alinhado com a agenda de inovação orientada por dados (*data-driven innovation*), promovendo o uso qualificado dos dados eleitorais para gerar valor social, educacional e político. Ao integrar ferramentas de exploração como APIs em SQL e R, o portal estimula a produção de conhecimento científico, a criação de aplicações derivadas e o uso pedagógico dos dados em contextos de formação cidadã. Em síntese, o Portal da Classe Política é mais do que um repositório técnico: é uma plataforma de inteligência cívica baseada em dados. Ao integrar o vasto acervo de informações eleitorais brasileiras e apresentá-lo de maneira clara, responsiva e confiável, a iniciativa promove um ciclo virtuoso de transparência, conhecimento e confiança. Nesse sentido, alinha-se a tendências globais de dados abertos e cidadania digital, consolidando-se como uma experiência inovadora e com potencial de impacto tanto nacional quanto internacional para o fortalecimento das instituições democráticas.

A Figura 1 faz uma representação gráfica dos objetivos do Portal da Classe Política e seus impactos esperados.

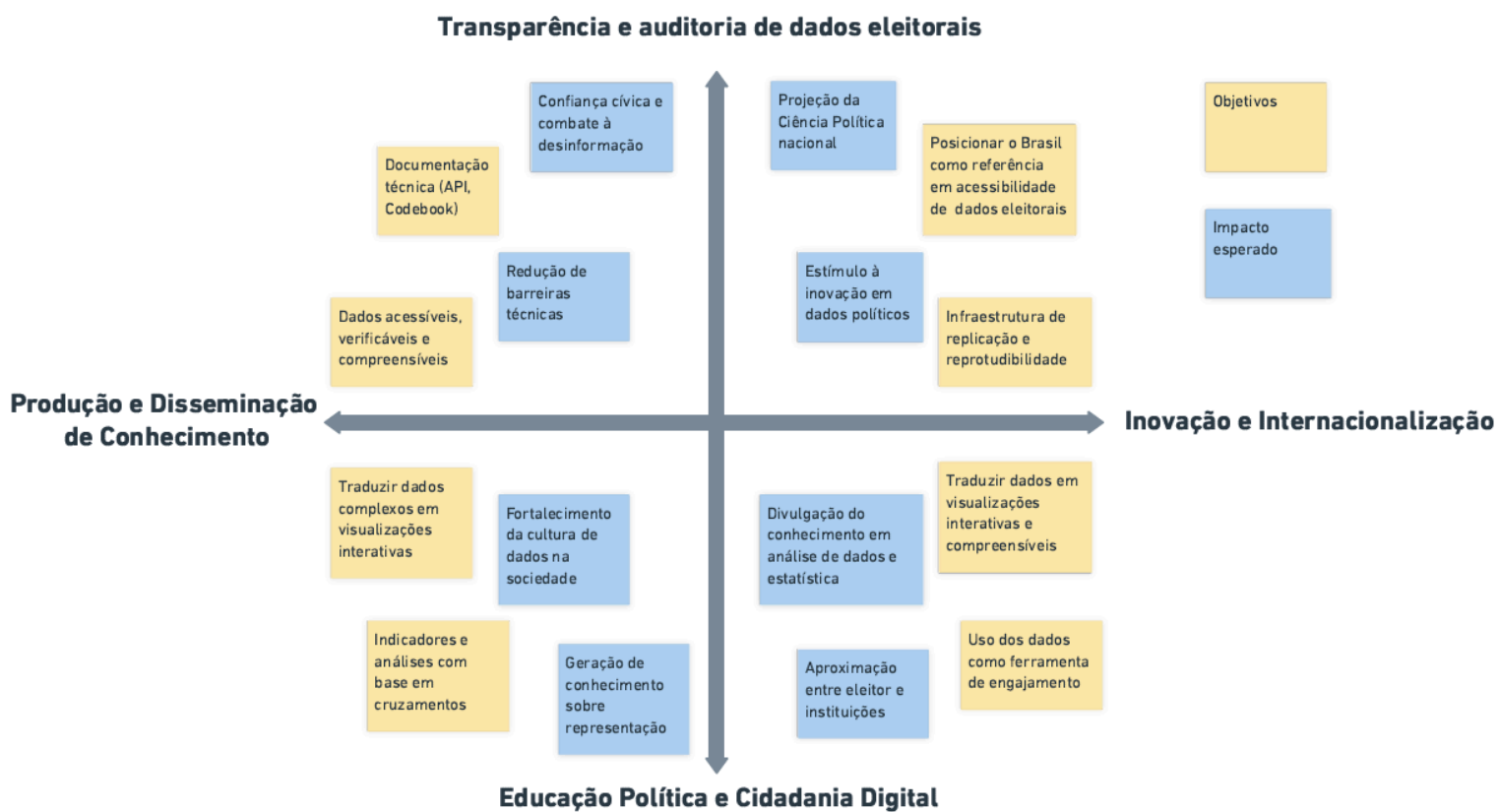


Figura 1 – Mapa de objetivos e impactos do Portal da Classe Política UFPR – INCT ReDem

Fonte: Elaboração própria (2024).

4. Descrição metodológica do projeto Portal da Classe Política

O Portal da Classe Política baseia-se integralmente em dados públicos fornecidos pelo Tribunal Superior Eleitoral (TSE) por meio de seu portal de dados abertos. O projeto abrange todas as eleições brasileiras de 1998 até os pleitos mais recentes, incluindo eleições gerais (presidenciais, legislativas estaduais e federais) e municipais (prefeitos, vice-prefeitos e vereadores). A escolha do marco temporal de 1998 deve-se à padronização dos bancos de dados implementada pelo TSE a partir desse ano, o que garantiu maior qualidade, completude e uniformidade na estrutura dos arquivos eleitorais.

Entre as principais fontes de dados utilizadas, destacam-se: (a) os registros de candidaturas, com informações detalhadas sobre o perfil dos candidatos (dados pessoais, filiação partidária, cargo disputado, situação da candidatura, entre outros); (b) os resultados de votação por seção eleitoral e município; (c) os dados de financiamento de campanha, contendo receitas declaradas, origens e doadores; e (d) as declarações de bens patrimoniais dos candidatos. Complementarmente, foram consolidados dados auxiliares como informações sobre quantidade de vagas por cargo, fundamentais para o cálculo de indicadores analíticos como o Quociente Partidário, Taxa de Reeleição, NEP, entre outros.

Os dados foram extraídos diretamente das fontes oficiais em formatos abertos (majoritariamente arquivos CSV). Apesar disso, seu uso direto é limitado por desafios técnicos: volumetria elevada, segmentação em centenas de planilhas e recorrentes problemas de padronização e formatação. Além disso, ao longo do desenvolvimento do projeto, foram identificadas lacunas de dados em algumas eleições, o que precisou de tratamento especial para a padronização e normalização dos bancos de dados. Por isso, o projeto estruturou um pipeline completo de coleta, tratamento, limpeza,

normalização e validação dos dados, resultando em um banco unificado e pronto para análises complexas.

A base consolidada foi implementada em um Sistema Gerenciador de Banco de Dados relacional (SGBD), dada a natureza altamente tabular dos dados e a necessidade de realizar cruzamentos entre diferentes dimensões (candidaturas, votações, financiamentos, bens etc.). O modelo de dados segue uma arquitetura entidade-relacionamento, com integridade referencial garantida entre tabelas factuais e dimensionais. Por exemplo, a tabela central de Candidaturas está interligada às tabelas de Votos, Financiamento e Patrimônio por meio de chaves estrangeiras comuns, permitindo a construção de consultas especializadas e indicadores derivados. Esse modelo é típico de *data warehouses* políticos, pensado para consultas rápidas, escalabilidade e manutenção modular.

Em termos de escala, o banco de dados estruturado no Portal da Classe Política agrega informações de 14 eleições entre 1998 e 2022, com um total de 2.884.495 candidaturas computadas, referentes a 1.822.832 indivíduos únicos.¹ O volume total de dados tratados atinge aproximadamente 8 GB em formato de *dataframe*, resultantes do processamento de 47 planilhas oficiais do TSE. Esses números expressam não apenas a complexidade técnica envolvida, mas também o compromisso do projeto com a organização, acessibilidade e inteligência sobre o sistema eleitoral brasileiro.

A abordagem de armazenamento adotada facilita atualizações e correções. À medida que novas eleições ocorrem, novas entradas podem ser acrescentadas às tabelas correspondentes sem necessidade de alterar o esquema. O banco de dados foi concebido para ser escalável, suportando consultas que englobam desde um único pleito até séries históricas completas. Destaca-se que, durante o desenvolvimento, foi criada uma instância de banco de dados principal para concentrar todas as informações integradas do projeto. Esse banco principal passou por processos de correção e atualização, integrando com sucesso as diversas planilhas oriundas do TSE.

Os dados coletados passam por um processo estruturado de padronização e enriquecimento antes de alimentar as análises disponíveis no portal. Inicialmente, cada conjunto de dados (candidaturas, votos, financiamento, bens) é harmonizado em

¹ O projeto é constantemente atualizado, conforme as atualizações de dados no TSE e em breve estará incorporado os dados eleitorais de 2024, além de novas funcionalidades e ajustes nos mecanismos atuais.

relação à nomenclatura das variáveis, formatação de códigos (municípios, partidos, unidades da federação) e unidades monetárias. Na etapa seguinte, os dados são integrados: registros de votação, receitas e bens são vinculados às respectivas candidaturas, preenchendo o banco relacional unificado.

Esse processo exigiu a definição de chaves de ligação — como o código SQ_CANDIDATO, utilizado pelo TSE para identificar candidaturas dentro de cada eleição — e, em diversos casos, a complementação de informações ausentes para assegurar o cruzamento correto entre tabelas. Durante o tratamento, inconsistências nas bases originais foram identificadas e corrigidas. Em 2002 e 2010, por exemplo, foi necessário ajustar a coluna SG_UE, que trazia códigos de municípios trocados. Já na eleição de 2000, detectou-se a ausência de 67 candidaturas na planilha oficial, problema resolvido com a inclusão de dados sobre candidatos substituídos por meio da variável SQ_SUBSTITUIDO.

Essas correções garantem a consistência interna do banco: cada voto está vinculado a um candidato válido, e cada candidato possui registros compatíveis de receitas e bens, evitando distorções nas análises. Com a base tratada e integrada foi desenvolvida uma camada de dados composta por indicadores analíticos derivados. Essa camada aplica fórmulas consagradas da Ciência Política e da Sociologia Política, automatizando cálculos que antes dependiam de tratamento manual.

Entre os principais indicadores calculados estão: o Número Efetivo de Partidos (NEP), que mede a fragmentação partidária; a Taxa de Renovação das casas legislativas, que compara a composição entre legislaturas; o Índice de Volatilidade Eleitoral de Pedersen, que resume mudanças no apoio partidário entre eleições, entre outros. Também são computadas métricas descritivas, como o número total de partidos com votos válidos ou com cadeiras conquistadas.

Ao todo, dezesseis indicadores foram definidos e implementados, abordando aspectos como competição partidária, volatilidade eleitoral e renovação das elites políticas. Cada indicador foi formalmente documentado, com especificação do nível de análise (nacional, estadual, municipal), filtros aplicáveis (por cargo, ano, região) e visualização de dados (gráficos de linha, barras e circular).

4.1 Processamento e tratamento de dados

O fluxo de dados inicia nos arquivos CSV brutos fornecidos por órgãos eleitorais, passa por etapas de leitura e tratamento no servidor *Node.js*, e termina com a inserção dos dados tratados no banco de dados. Em termos simples, o sistema lê os arquivos, valida e organiza as informações, e armazena tudo de forma consistente no banco, garantindo que diferentes conjuntos de dados (candidatos, eleições, votos, etc.) fiquem interligados. O objetivo geral do sistema é automatizar esse processo de ETL (Extração, Transformação e Carga) de forma confiável, mesmo para um público não especializado entender como os dados são processados.

O *back-end* foi implementado em *JavaScript*, executando em ambiente *Node.js*, aproveitando sua capacidade de lidar com operações de I/O (entrada/saída) de forma eficiente. Foi utilizado o framework *Sequelize*, um ORM (*Object-Relational Mapping*) para *Node.js*, que simplifica a interação com o banco de dados relacional através de modelos de dados em *JavaScript*. Isso permite manipular tabelas do banco como objetos e realizar consultas e inserções com métodos da linguagem, sem escrever SQL manualmente na maioria dos casos. Na Tabela 1, estão descritas as bibliotecas utilizadas no processamento dos dados e no desenvolvimento *back-end*.

Tabela 1 – Bibliotecas em *JavaScript* utilizadas na arquitetura do projeto

Biblioteca	Descrição resumida
Csv-parser	Realiza a leitura de arquivos CSV linha a linha, convertendo cada registro em objetos JavaScript.
Moment	Manipula e formata datas, convertendo strings como DD/MM/YYYY em objetos Date do JavaScript.
Json2csv/csv-writer	Gera arquivos CSV a partir de objetos JavaScript, usados para logs de erro ou reprocessamento de dados.
Fast-csv	Utilitário para escrita rápida de CSVs, empregado na exportação otimizada de grandes volumes de dados

Iconv-lite	Converte arquivos com encoding ISO-8859-1 (Latin1) para UTF-8, garantindo a leitura correta de acentos.
RabbitMQ (amqplib)	Permite o uso de filas para processar grandes volumes de dados de forma assíncrona, evitando sobrecarga do banco.

Fonte: Elaboração própria (2024).

O banco de dados em uso é relacional (através do *PostgreSQL*), organizado em tabelas que refletem entidades do domínio eleitoral. A modelagem inclui, entre outros:

- Candidato: informações pessoais básicas dos candidatos (nome, data de nascimento, município de nascimento etc.), além de identificadores como CPF e título de eleitor.
- Eleição: dados de cada eleição (ano, turno, tipo de eleição).
- Cargo: cargos disputados (por exemplo, Presidente, Governador, Prefeito).
- Partido: partidos políticos com suas siglas e nomes.
- Candidato Eleição: tabela associativa que liga um candidato a uma eleição específica, armazenando dados da candidatura naquele pleito (número sequencial do candidato, situação da candidatura, coligação, votação, etc.).
- Unidade Eleitoral: unidade da federação ou região eleitoral (por exemplo, estados, município ou código "BR" para eleições nacionais) que contextualiza a eleição ou candidatura.
- Ocupação: profissão ou ocupação declarada do candidato.
- Grau de Instrucao: nível de instrução do candidato.
- Gênero e Raça: gênero (masculino, feminino) e raça/cor autodeclarada do candidato.
- Situação Candidatura e Situação Turno: status da candidatura (deferida, indeferida etc.) e resultado no turno (eleito, não eleito, suplente, etc.).
- Nome Urna: nome do candidato conforme exibido na urna eletrônica, que às vezes difere do nome civil.
- Situação Reeleição: indica se o candidato tentava reeleição.

- Municípios Votação e Votação Candidato Município: tabelas para os resultados de votação, ligando cada candidato em uma eleição aos votos que ele recebeu em cada município.
- Categoria Bens e Bens Candidato Eleição: categorias de bens declarados (imóveis, veículos etc. com um código identificador) e os bens declarados por cada candidato na eleição (descrição do bem e valor declarado).
- Doador e Doações Candidato Eleição: informações de doadores (CPF/CNPJ, nome) e registro de doações de campanha recebidas por um candidato em determinada eleição, incluindo natureza da receita e valor.

Essa estrutura de modelos assegura que os dados importados dos CSV sejam armazenados de forma normalizada. Por exemplo, informações repetidas em várias linhas do CSV (como o partido de vários candidatos) são armazenadas uma única vez na tabela de Partido e referenciadas via chave estrangeira, em vez de texto repetido em cada registro de candidato. Isso economiza espaço e mantém a consistência (se o nome de um partido precisar ser corrigido, ajusta-se em um lugar só).

Ao final, o desenvolvimento *back-end* resultou em um pipeline robusto para carregar dados eleitorais históricos no banco de dados, mantendo a integridade e oferecendo rastreabilidade. Foram enfrentados diversos desafios, entre eles: lidar com arquivos grandes e de *encoding* variado, consolidar entidades que aparecem em diferentes fontes (por exemplo, relacionar um candidato às suas candidaturas e votações), e garantir desempenho ao inserir milhões de registros. A solução encontrou um equilíbrio através de técnicas como processamento em lotes paralelos, uso de ORM para simplificar operações e evitar erros de SQL, e verificação cuidadosa de pré-requisitos de cada inserção. Um dos maiores aprendizados foi a importância de tratar a qualidade dos dados de entrada: muitas regras de normalização e verificação tiveram que ser implementadas para adequar os dados brutos à estrutura do sistema, o que reforça a necessidade de uma etapa de transformação bem planejada em projetos ETL. Os logs detalhados produzidos também se mostraram essenciais para iterar no processo – permitiram identificar pontos fracos nos dados e ajustar o pipeline ou os próprios dados para a próxima execução.

5. Aplicações práticas do Portal da Classe Política

Uma das funcionalidades centrais do Portal da Classe Política é o cruzamento interativo de indicadores eleitorais e partidários, disponíveis na aba “Eleitorais e Partidários”. Essa seção permite ao usuário explorar métricas consolidadas sobre o funcionamento do sistema político, utilizando filtros como período, cargo, unidade federativa e indicador desejado.

No exemplo apresentado na Imagem 2, o usuário seleciona o indicador Número Efetivo de Partidos (NEP - Legislativo) para o cargo de Deputado Federal, considerando todos os estados brasileiros no período de 1998 a 2022. O NEP é uma métrica consagrada na Ciência Política que mensura o grau de fragmentação partidária em uma determinada eleição, levando em conta não apenas o número de partidos com representação, mas também o peso relativo de cada um deles.

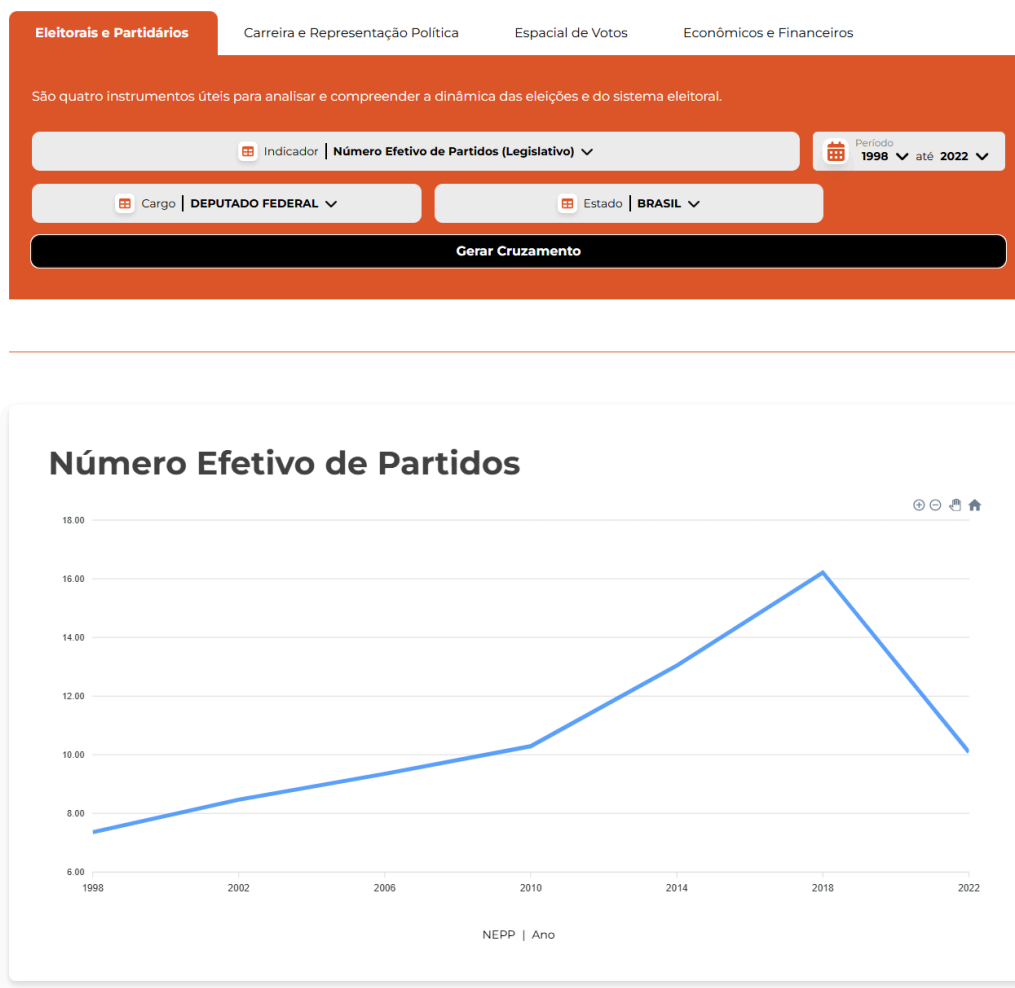
A funcionalidade de indicadores eleitorais do Portal da Classe Política permite acompanhar a evolução do Número Efetivo de Partidos (NEP) na Câmara dos Deputados entre 1998 e 2022. O gráfico apresentado revela uma tendência de crescimento constante na fragmentação partidária ao longo das últimas décadas, com um pico acentuado em 2018, seguido de uma queda significativa em 2022.

Em 1998, o NEP girava em torno de 7,5, refletindo um sistema ainda relativamente concentrado. Ao longo dos anos, o número foi crescendo: ultrapassou 10 partidos efetivos em 2010, atingiu aproximadamente 12,5 em 2014 e chegou ao seu ponto mais alto em 2018, com cerca de 16 legendas efetivas compondo a Câmara dos Deputados. Esse valor expressa o auge da dispersão partidária, em um contexto de baixa governabilidade e alta fragmentação.

A partir de 2022, observa-se uma queda expressiva do NEP, que recua para aproximadamente 10,5. Essa inflexão sugere os primeiros efeitos concretos das reformas institucionais recentes, como a cláusula de desempenho (barreira) e o fim das coligações proporcionais, que passaram a impactar diretamente a sobrevivência e competitividade de pequenos partidos. A curva descendente pode indicar um novo ciclo de racionalização do sistema partidário.

Essa análise, viabilizada com poucos cliques na interface do portal, ilustra como os dados abertos podem ser transformados em conhecimento acessível e acionável. A ferramenta permite ainda desagregar os dados por estado, município ou cargo e comparar o resultado das legislaturas eleitas. Em constante evolução, as análises estão sendo incrementadas com textos explicativos que sintetizam o que significam os índices e indicadores, tornando ainda mais didática a visualização dos dados e a escolha do indicador.

Figura 2 – Simulação de usuário ao analisar o Número Efetivo de Partidos no Portal



da Classe Política

Fonte: Elaboração própria (2024).

Na aba "Econômicos e Financeiros", o Portal da Classe Política disponibiliza ferramentas para explorar a dimensão patrimonial e financeira das campanhas eleitorais. Um dos indicadores é a Média e Mediana de Patrimônio da Classe Política, que permite acompanhar como evoluem os bens declarados pelos candidatos ao longo do tempo.

No exemplo apresentado na Imagem 3, foi selecionado o cargo de Deputado Federal no período de 2006 a 2022, abrangendo cinco eleições gerais consecutivas. Os dados revelam um comportamento interessante: a média de patrimônio dos candidatos, representada pela linha azul, apresenta flutuações significativas, enquanto

a mediana, mostrada em verde, segue uma tendência de crescimento mais linear e estável.

Em 2006, a média patrimonial era próxima de R\$ 300 mil, mas caiu em 2010, mantendo-se praticamente estável em 2014. A partir de 2018, há um salto expressivo, atingindo quase R\$ 400 mil, seguido de uma queda novamente em 2022, com média próxima a R\$ 250 mil. A mediana, por outro lado, cresce de forma constante, saindo de R\$ 20 mil em 2006 para cerca de R\$ 60 mil em 2022.

Essa discrepância entre média e mediana é reveladora: enquanto a mediana representa o valor mais típico entre os candidatos, a média é puxada para cima por candidaturas com patrimônio muito elevado. O gráfico sugere que, mesmo com variações pontuais, a elite política brasileira permanece altamente concentrada em termos de renda e patrimônio, com uma presença persistente de candidatos milionários.

O portal ainda permite aprofundar essa análise por estado, ano ou cargo, possibilitando comparações regionais e por ciclo eleitoral. Para pesquisadores e jornalistas, esse tipo de cruzamento permite identificar padrões de concentração econômica, avaliar o perfil financeiro dos eleitos e investigar eventuais distorções no acesso à política.

Figura 3 – Simulação de usuário ao analisar a média e mediana de patrimônio no Portal da Classe Política

Eleitorais e Partidários Carreira e Representação Política Espacial de Votos **Econômicos e Financeiros**

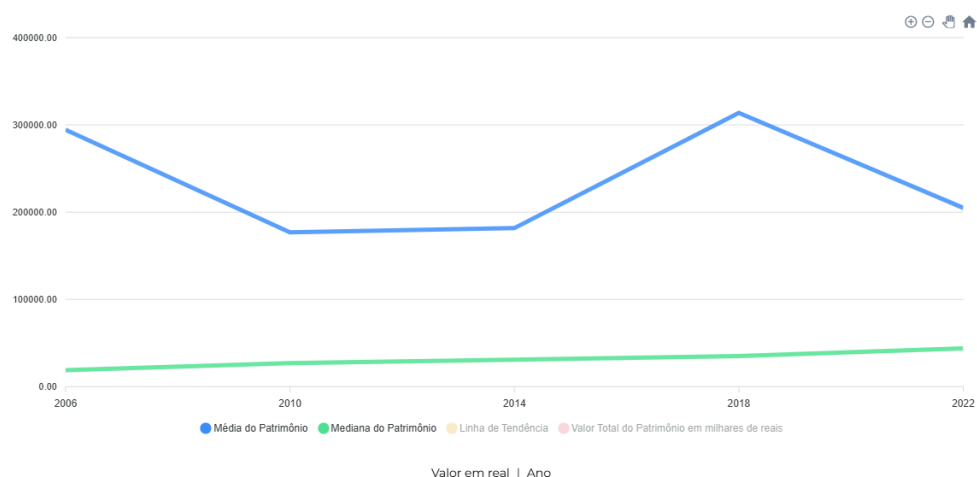
São quatro instrumentos úteis para analisar e compreender as dinâmicas econômica e financeira das campanhas eleitorais.

Indicador | **Média e Mediana de Patrimônio da Classe Política** Período | 2006 ▼ até 2022 ▼

Cargo | **DEPUTADO FEDERAL** ▼ Estado | **BRASIL** ▼

Gerar Cruzamento

Média e Mediana de Patrimônio da Classe Política



Fonte: Elaboração própria (2024).

O Portal da Classe Política permite ao usuário realizar buscas diretas por indivíduos que tenham concorrido a cargos eletivos desde 1998. A ferramenta de “Consulta por Candidato” oferece uma visão integrada da trajetória política e eleitoral de qualquer figura pública registrada no sistema do TSE. Esse recurso é especialmente útil para jornalistas, eleitores, pesquisadores e organizações da sociedade civil interessadas em avaliar a consistência, o histórico e o perfil socioeconômico de seus representantes.

No exemplo apresentado na Figura 4, a busca é feita pelo nome “Gleisi”, restringida às eleições gerais no estado do Paraná. O portal retorna o resultado de forma imediata, apresentando o nome completo, partido, cargo, situação da candidatura e um botão de “Mais Informações” que leva à página de perfil detalhado da candidata.

Na tela seguinte, o usuário encontra um conjunto completo de informações da última eleição disputada. Isso inclui: Dados básicos: nome completo, idade, gênero, cor/raça, escolaridade, ocupação e cidade de nascimento; Informações político-eleitorais: partido, coligação, estado de candidatura e situação da candidatura (apto/inapto); Dados econômicos: valor declarado de bens patrimoniais.

Em seguida, o portal apresenta um histórico de votações por eleição, permitindo observar a evolução do desempenho eleitoral da candidata ao longo do tempo. No caso de Gleisi Hoffmann, é possível identificar sua participação em pleitos desde 2006 até 2022, ficando de fora das disputas apenas nas municipais de 2012 e 2016.

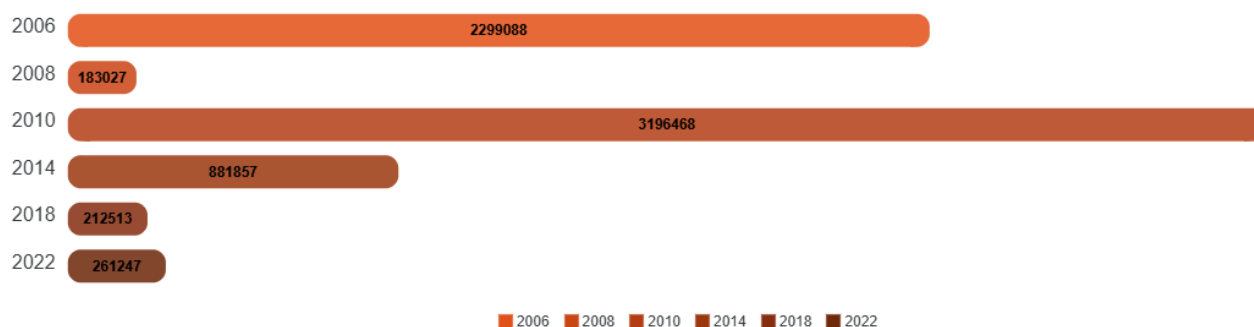
Complementando a análise, o sistema também oferece um mapa da votação na última eleição, com os votos recebidos por município. Essa visualização espacial revela onde o candidato tem maior força eleitoral dentro do estado, o que é essencial para estudos sobre base territorial, enraizamento político ou padrões regionais de votação.

A consulta individual sintetiza múltiplos conjuntos de dados — candidaturas, bens, resultados, coligações e votações — em uma interface única e acessível. Trata-se de uma ferramenta para promover accountability, transparência e engajamento cívico, possibilitando ao cidadão comum avaliar de forma crítica e fundamentada quem são os representantes que compõem a classe política brasileira.

Figura 4 – Simulação de usuário ao consultar uma candidatura no Portal da Classe Política



Histórico de Votações



Fonte: Elaboração própria (2024).

6. Considerações

O Portal da Classe Política representa uma iniciativa de inovação pública com forte impacto sobre duas dimensões estratégicas da vida democrática: a transparência eleitoral e o avanço da Ciência Política empírica no Brasil. Ao transformar grandes volumes de dados brutos da Justiça Eleitoral em visualizações acessíveis, indicadores analíticos e ferramentas interativas, o projeto promove o acesso qualificado à informação política e eleitoral, contribuindo diretamente para o fortalecimento da *accountability*, da confiança nas instituições e da cultura cívica baseada em evidências.

Para a comunidade acadêmica, o portal se consolida como um recurso técnico de alto valor, ampliando as possibilidades de pesquisa empírica com dados eleitorais padronizados, auditáveis e organizados em uma estrutura relacional robusta. Os indicadores disponibilizados, como número efetivo de partidos, taxa de reeleição, volatilidade eleitoral, entre outros, permitem análises replicáveis e comparações históricas com base em fundamentos da Ciência Política. A iniciativa também avança sobre a fronteira da ciência aberta, ao democratizar o acesso a bancos de dados normalmente restritos a usuários experientes em programação ou ciência de dados.

O projeto é mantido por uma equipe multidisciplinar composta por cientistas políticos, analistas de dados e desenvolvedores, que atua de forma contínua na manutenção, validação e expansão do sistema. Trata-se de uma plataforma viva, que está em constante aperfeiçoamento técnico e metodológico, com o objetivo de incorporar novas funcionalidades, corrigir inconsistências e responder às demandas de seus públicos diversos — pesquisadores, jornalistas, profissionais da política, estudantes, gestores públicos e cidadãos engajados.

Entre as etapas em curso, destaca-se o desenvolvimento de uma API pública para o portal, que permitirá aos usuários consultar diretamente o banco de dados relacional e extrair os resultados dos indicadores calculados, inclusive por meio de requisições parametrizadas. Essa infraestrutura ampliará a interoperabilidade dos dados e facilitará sua integração com aplicações externas, visualizações personalizadas e análises automatizadas.

Além disso, está previsto o lançamento de um pacote em linguagem R que fornecerá funções específicas para acessar, filtrar e explorar os dados disponíveis no Portal da Classe Política, diretamente do ambiente de análise estatística. Esse recurso facilitará ainda mais o uso dos dados por pesquisadores, professores e estudantes familiarizados com o universo da ciência de dados aplicada à Ciência Política.

Com esses desdobramentos, o Portal da Classe Política consolida seu compromisso com a produção de conhecimento aberto, qualificado e socialmente relevante, aproximando a pesquisa acadêmica da sociedade e fortalecendo as bases para uma democracia mais informada, participativa e confiável. Trata-se de um projeto em construção permanente, que olha para o futuro com foco na expansão do acesso à informação, na sofisticação das análises eleitorais e na consolidação de uma infraestrutura pública de dados à altura dos desafios do século XXI.

7. Referências bibliográficas

- Bezerra Da Silva, M., Fernandes De Barros Costa Azevedo, R., Oliveira De Araújo, D., Percia França, F., & Da Silva Pereira, M. (2020). A publicidade de dados abertos pelo Tribunal Superior Eleitoral (TSE): O caso do Repositório de Dados Eleitorais. *Ciência da Informação*, 49(3). <https://doi.org/10.18225/ci.inf.v49i3.5228>
- Brandt, M. B., Vidotti, S. A. B. G., & Segundo, J. E. S. (2018). Modelo de dados abertos conectados para informação legislativa. *Informação & Sociedade*, 28(2), Artigo 2. <https://periodicos.ufpb.br/index.php/ies/article/view/37979>
- Castro, J. (2022, maio 11). *Pesquisa aponta que 40% dos brasileiros confiam muito nas urnas eletrônicas*. JOTA Jornalismo. <https://www.jota.info/eleicoes/pesquisa-aponta-que-40-dos-brasileiros-confiam-muito-nas-urnas-eletronicas>
- Duenas-Cid, D., Loeber, L., Martin-Rozumilowicz, B., & Macias, R. (2024). Trust and electoral technologies throughout the election cycle: Comparing the USA, Netherlands, Poland, and Kenya. *JeDEM - eJournal of eDemocracy and Open Government*, 16(3). <https://doi.org/10.29379/jedem.v16i3.922>
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*. <https://www.tandfonline.com/doi/abs/10.1080/10580530.2012.716740>
- Kaplan, C. A. (2006). *A guide to transparency in election administration*. IFES Consultant.

Quase 90% dos brasileiros admitem ter acreditado em fake news. (2024, abril 1).

Agência Brasil.

<https://agenciabrasil.ebc.com.br/geral/noticia/2024-04/quase-90-dos-brasileiros-admitem-ter-acreditado-em-fake-news>

Ruijter, E., Détienne, F., Baker, M., Groff, J., & Meijer, A. J. (2020). The Politics of Open Government Data: Understanding Organizational Responses to Pressure for More Transparency. *The American Review of Public Administration*, 50(3), 260–274.

<https://doi.org/10.1177/0275074019888065>