

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
INSTITUTO MULTIDISCIPLINAR

DAVID AUGUSTO FERNANDES FILHO  
JOÃO PEDRO FELIX LIMA

**Rural.Data: Portal de Dados e  
Transparência da UFRRJ**

Prof. Filipe Braida do Carmo, D.Sc.  
Orientador

Nova Iguaçu, Dezembro de 2025

# **Rural.Data: Portal de Dados e Transparência da UFRRJ**

**David Augusto Fernandes Filho**

**João Pedro Felix Lima**

Projeto Final de Curso submetido ao Departamento de Ciência da Computação do Instituto Multidisciplinar da Universidade Federal Rural do Rio de Janeiro como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Apresentado por:

---

David Augusto Fernandes Filho

---

João Pedro Felix Lima

Aprovado por:

---

Prof. Filipe Braida do Carmo, D.Sc.

---

Prof. Marcel William Rocha da Silva, D.Sc.

---

Prof. Juliana Mendes Nascente e Silva Zamith, D.Sc.

NOVA IGUAÇU, RJ - BRASIL

Dezembro de 2025



**DOCUMENTOS COMPROBATÓRIOS Nº 33248/2025 - CoordCGCC (12.28.01.00.00.98)**

*(Nº do Protocolo: NÃO PROTOCOLADO)*

**(Assinado digitalmente em 16/12/2025 08:53 )**

FILIPE BRAIDA DO CARMO  
PROFESSOR DO MAGISTERIO SUPERIOR  
DeptCC/IM (12.28.01.00.00.83)  
Matrícula: ###295#4

**(Assinado digitalmente em 16/12/2025 09:22 )**

JULIANA MENDES NASCENTE E SILVA ZAMITH  
COORDENADOR CURS/POS-GRADUACAO - TITULAR  
CoordCGCC (12.28.01.00.00.98)  
Matrícula: ###731#0

**(Assinado digitalmente em 16/12/2025 08:13 )**

MARCEL WILLIAM ROCHA DA SILVA  
PROFESSOR DO MAGISTERIO SUPERIOR  
PPGIHD (11.39.00.16)  
Matrícula: ###807#6

**(Assinado digitalmente em 16/12/2025 15:05 )**

JOAO PEDRO FELIX LIMA  
DISCENTE  
Matrícula: 2021#####4

**(Assinado digitalmente em 16/12/2025 14:58 )**

DAVID AUGUSTO FERNANDES FILHO  
DISCENTE  
Matrícula: 2021#####9

Visualize o documento original em <https://sipac.ufrrj.br/documentos/> informando seu número: 33248, ano: 2025, tipo: DOCUMENTOS COMPROBATÓRIOS, data de emissão: 15/12/2025 e o código de verificação: c6d6b07f32

# Agradecimentos

David Augusto Fernandes Filho

Primeiramente, agradeço à minha família, em especial aos meus pais. Se não fosse por eles, eu não estaria concluindo esta graduação. Foram eles que me educaram e me transmitiram suas visões de mundo, muitas vezes a partir de perspectivas bastante diferentes, o que me ajudou a enxergar a realidade de forma mais plural e a me tornar quem sou hoje.

Agradeço também aos meus animais de estimação, que estiveram ao meu lado em toda essa jornada — Laika, Morena, Pandora, Chimbica e Amora. Mesmo sem compreenderem o motivo do estresse dos seres humanos, foram fundamentais em diversos momentos, oferecendo sua presença e amor incondicional.

Registro, ainda, meu agradecimento a todos os professores e membros do Departamento de Ciência da Computação da Universidade Federal Rural do Rio de Janeiro (UFRRJ-IM). Em especial, ao meu orientador, Prof. Filipe Braida, pela paciência, dedicação e disponibilidade. Além de me guiar na construção deste trabalho, sempre reservou tempo para ouvir meus questionamentos após as aulas, sem pressa. É evidente o quanto gosta do que faz, e isso certamente contribuiu de forma significativa para a minha formação acadêmica e profissional.

Por fim, agradeço aos amigos da faculdade que compartilharam comigo inúmeras situações ao longo do curso. Entre risadas e reclamações, tornaram a experiência da graduação muito mais leve e especial.

João Pedro Felix Lima

Primeiramente, agradeço a Deus, por ter me dado saúde e força para superar os desafios desta jornada e concluir esta etapa tão importante da minha vida.

Aos meus pais, por todo o suporte, incentivo e pelos sacrifícios feitos para que eu pudesse ter acesso a uma educação de qualidade. Vocês são a base de tudo o que conquistei até aqui.

Ao meu orientador, Prof. Filipe Braida do Carmo, pela paciência, pelas diretrizes técnicas e por conduzir este trabalho com sabedoria. Obrigado por compartilhar seu conhecimento e contribuir para o meu crescimento acadêmico e profissional.

Aos amigos que fiz na UFRRJ, que tornaram os dias de estudo mais leves e os momentos de pressão mais suportáveis. Agradeço pela parceria, pelas risadas nos intervalos e pelo apoio mútuo durante as disciplinas e projetos.

E um agradecimento mais do que especial à minha companheira leal, minha gata Kimmy. Obrigado por estar sempre ali, "sentada" ao meu lado (ou em cima do teclado) durante as longas madrugadas de código e escrita. Obrigado por pedir carinho exatamente nos momentos em que eu precisava de uma pausa e por "ouvir" pacientemente todas as minhas reclamações quando o sistema não funcionava como esperado. Sua presença silenciosa e reconfortante foi fundamental para manter minha sanidade e alegria neste processo.

## RESUMO

Rural.Data: Portal de Dados e Transparência da UFRRJ

David Augusto Fernandes Filho e João Pedro Felix Lima

Dezembro/2025

Orientador: Filipe Braida do Carmo, D.Sc.

A transparência ativa na administração pública é um pilar essencial para o controle social, contudo, sua efetividade é frequentemente comprometida pela dispersão de dados e pela predominância de formatos estáticos, que dificultam a análise aprofundada e o acesso à informação pela sociedade. Diante desse cenário, este trabalho apresenta o desenvolvimento do Rural.Data, um portal de dados e transparência concebido para a Universidade Federal Rural do Rio de Janeiro (UFRRJ). A proposta consiste na implementação de uma plataforma web robusta que integra um módulo de ingestão dinâmica, capaz de processar arquivos de dados com estruturas heterogêneas através de um fluxo automatizado de ETL (*Extract, Transform, Load*), garantindo a integridade e padronização das informações. Adicionalmente, o sistema incorpora uma camada de inteligência de negócios automatizada, orquestrando a ferramenta Metabase para a geração instantânea de painéis visuais interativos. Diferente de soluções convencionais que atuam meramente como repositórios de arquivos, o Rural.Data entrega uma experiência de visualização analítica e atualizável, validando a viabilidade técnica de uma infraestrutura moderna para democratizar o acesso aos indicadores institucionais e otimizar a gestão da informação pública.

## ABSTRACT

Rural.Data: Portal de Dados e Transparência da UFRRJ

David Augusto Fernandes Filho and João Pedro Felix Lima

Dezembro/2025

Advisor: Filipe Braida do Carmo, D.Sc.

*Active transparency in public administration is an essential pillar for social control; however, its effectiveness is often compromised by data dispersion and the predominance of static formats, which hinder in-depth analysis and access to information by society. Given this scenario, this work presents the development of Rural.Data, a data and transparency portal designed for the Federal Rural University of Rio de Janeiro (UFRRJ). The proposal consists of implementing a robust web platform that integrates a dynamic ingestion module capable of processing data files with heterogeneous structures through an automated ETL (Extract, Transform, Load) workflow, ensuring information integrity and standardization. Additionally, the system incorporates an automated business intelligence layer, orchestrating the Metabase tool for the instant generation of interactive visual dashboards. Unlike conventional solutions that act merely as file repositories, Rural.Data delivers an analytical and updatable visualization experience, validating the technical feasibility of a modern infrastructure to democratize access to institutional indicators and optimize public information management.*

# List a de Figuras

Figura 2.1: Arquitetura simplificada de solução de <i>Business Intelligence</i> baseada em processo de <i>Extract, Transform, Load</i> (ETL) e <i>data warehouse</i> . . . . .	8
Figura 2.2: Exemplo de ecossistema analítico com <i>data lake</i> , <i>data warehouses</i> e camadas de análise e BI. . . . .	9
Figura 2.3: Visão esquemática do processo de ETL. . . . .	11
Figura 3.1: Exemplo de catálogo de conjuntos de dados em portal institucional de dados abertos. . . . .	20
Figura 3.2: DER do banco de dados do portal de dados e transparência proposto. .	35
Figura 4.1: Estrutura de diretórios do módulo de <i>Business Intelligence</i> . . . . .	43
Figura 4.2: Arquitetura em camadas adotada no sistema. . . . .	44
Figura 4.3: Interface pública de visualização de painel de indicadores. . . . .	47
Figura 4.4: Visualização detalhada de um painel de indicadores. . . . .	47
Figura 4.5: Tela de cadastro de usuário. . . . .	49
Figura 4.6: Tela de login do sistema. . . . .	50
Figura 4.7: Painel de gestão de conjuntos de dados ( <i>datasets</i> ). . . . .	51
Figura 4.8: Formulário de importação de novo conjunto de dados. . . . .	52

Figura 4.9: Interface de controle de painéis vinculados a um *dataset*. . . . . 53

# Lista de Abreviaturas e Siglas

**MVC** *Model View Controller*

**CLI** *Command Line Interface*

**SPA** *Single Page Application*

**HTML** *HyperText Markup Language*

**API** *Application Programming Interface*

**SGBD** Sistema Gerenciador de Banco de Dados

**SQL** *Structured Query Language*

**ACID** Atomicidade, Consistência, Isolamento e Durabilidade

**BI** *Business Intelligence*

**ETL** *Extract, Transform, Load*

**CSV** *Comma Separated Values*

**JSON** *JavaScript Object Notation*

**ORM** *Object-Relational Mapping*

**UUID** *Universally Unique Identifier*

**IFES** Instituições Federais de Ensino Superior

**PDA** Plano de Dados Abertos

**UFRRJ** Universidade Federal Rural do Rio de Janeiro

**SSO** *Single Sign-On*

# Sumário

<b>Agradecimentos</b>	i
<b>Resumo</b>	iii
<b>Abstract</b>	iv
<b>Lista de Figuras</b>	v
<b>Lista de Abreviaturas e Siglas</b>	vii
<b>1 Introdução</b>	1
1.1 Objetivo . . . . .	3
1.2 Organização do Trabalho . . . . .	4
<b>2 Análise de dados e Business Intelligence</b>	6
2.1 Conceitos de <i>Business Intelligence</i> . . . . .	6
2.2 Arquiteturas analíticas e modelagem de dados para BI . . . . .	7
2.3 Processos de ETL em ambientes de BI . . . . .	10
2.4 <i>Business Intelligence</i> na gestão universitária . . . . .	12
2.5 Síntese da fundamentação teórica . . . . .	12

<b>3 Proposta</b>	<b>14</b>
3.1 Motivação . . . . .	15
3.2 Trabalhos relacionados . . . . .	18
3.3 Especificação da solução . . . . .	21
3.3.1 Visão geral do desenho do sistema . . . . .	21
3.3.2 Decisões arquiteturais da proposta . . . . .	23
3.3.3 Atores e perfis de uso . . . . .	24
3.3.4 Requisitos funcionais . . . . .	25
3.3.5 Requisitos não funcionais . . . . .	27
3.3.6 Regras de negócio . . . . .	28
3.3.7 Casos de uso e fluxos principais . . . . .	29
3.4 Banco de dados . . . . .	34
3.4.1 Diagrama entidade–relacionamento . . . . .	34
<b>4 Rural.Data: Portal de Dados e Transparência da UFRRJ</b>	<b>37</b>
4.1 Tecnologias Utilizadas . . . . .	37
4.1.1 AdonisJS . . . . .	38
4.1.2 React . . . . .	38
4.1.3 Inertia.js . . . . .	39
4.1.4 Tailwind CSS . . . . .	39
4.1.5 ShadCN/UI . . . . .	40
4.1.6 PostgreSQL . . . . .	40
4.1.7 Docker . . . . .	40

4.1.8	Metabase	41
4.2	Implementação do Sistema	42
4.2.1	Estrutura de Arquivos e Organização	42
4.2.2	Arquitetura em Camadas	43
4.2.2.1	Camada de Apresentação	44
4.2.2.2	Camada de Lógica de Negócio	44
4.2.2.3	Camada de Acesso a Dados	45
4.2.3	Portal Público e Visualização de Dados	46
4.2.4	Sistema de Autenticação e Controle de Acesso	48
4.2.5	Gerenciamento de Conjuntos de Dados e Painéis	50
<b>5</b>	<b>Considerações Finais</b>	<b>54</b>
5.1	Síntese do trabalho desenvolvido	54
5.2	Contribuições do trabalho	56
5.3	Limitações e trabalhos futuros	57
5.4	Considerações finais	59
<b>Referências</b>		<b>60</b>

# Capítulo 1

## Introdução

A expansão do volume de dados produzidos por organizações públicas e privadas nas últimas décadas tornou a informação um recurso estratégico para a tomada de decisão, o planejamento e a avaliação de políticas. No setor público, esse movimento é acompanhado por iniciativas de transparência e participação social que buscam disponibilizar dados governamentais em formatos abertos, reutilizáveis e acessíveis à sociedade. Nesse contexto, a pauta de *dados abertos governamentais* ganha relevância como elemento central de agendas de governo aberto, inovação em serviços públicos e controle social.

No Brasil, a Política de Dados Abertos do Poder Executivo federal, instituída pelo Decreto nº 8.777/2016, estabelece diretrizes para a publicação de bases de dados em portais unificados, privilegiando formatos padronizados e a atualização contínua das informações Brasil (2016). O Portal Brasileiro de Dados Abertos<sup>1</sup> passou a centralizar grande parte desses *datasets*, reforçando a importância de catálogos públicos como instrumentos para ampliar a transparência ativa e fomentar o desenvolvimento de soluções baseadas em dados.

Esse movimento também alcança as Instituições Federais de Ensino Superior (IFES), que produzem e mantêm um conjunto expressivo de dados acadêmicos, administrativos, orçamentários e de pesquisa. Estudos recentes indicam, entretanto, que

---

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

a oferta de dados abertos pelas universidades federais brasileiras ainda é marcada por heterogeneidade quanto ao grau de maturidade e à articulação entre planejamento, publicação e uso efetivo das informações. Albano, Craveiro e Gaffrée (2021) mostram que o Plano de Dados Abertos (PDA) das IFES avançam na formalização de compromissos, mas enfrentam desafios em governança, priorização e atualização de conjuntos de dados. Na mesma direção, Pires (2019) ressalta que o envolvimento interno de gestores e comunidades acadêmicas é decisivo para que os dados publicados extrapolem o cumprimento formal de normas e passem a sustentar práticas de transparência e prestação de contas à sociedade. Já Lima e Pires (2022) destacam dificuldades de acesso, fragmentação e limitações de usabilidade em parte dos portais hoje existentes.

No âmbito da gestão universitária, a simples disponibilização de arquivos para *download* em portais de dados abertos não garante, por si só, a apropriação estratégica da informação. Por um lado, gestores, unidades acadêmicas e órgãos de planejamento carecem de visões consolidadas, indicadores e análises que apoiem decisões em áreas como orçamento, assistência estudantil, evasão, permanência e desempenho acadêmico. Por outro, a comunidade externa, composta por estudantes, pesquisadores, jornalistas e cidadãos, frequentemente se depara com barreiras técnicas para explorar arquivos desestruturados, dispersos em múltiplas páginas e pouco integrados a ferramentas analíticas.

Ao mesmo tempo, conceitos e práticas de *Business Intelligence* (BI) consolidaram-se como alternativas para transformar grandes volumes de dados em informações relevantes para o suporte à decisão (LOSHIN, 2012; TURBAN; SHARDA; DELEN, 2014). Em arquiteturas típicas de BI, processos de ETL alimentam repositórios analíticos, como *data warehouses* e *data marts*, a partir dos quais *dashboards* e relatórios são construídos Inmon (2005), Kimball e Ross (2013). Na esfera acadêmica, aplicações de BI para apoio à gestão universitária têm demonstrado potencial para organizar e monitorar dados de produção científica, desempenho discente e processos administrativos (JÚNIOR et al., 2022; VIEIRA; SEABRA; GARCÉS, 2025).

Diante desse cenário, observa-se uma lacuna importante: enquanto portais institu-

cionais de dados abertos tendem a privilegiar a catalogação e o *download* de arquivos, muitas das soluções de BI permanecem restritas a ambientes internos, voltados apenas a gestores e equipes técnicas. Falta, portanto, uma camada intermediária que articule essas duas dimensões, permitindo que a publicação de dados abertos dialogue diretamente com painéis analíticos, visualizações interativas e narrativas de transparência compreensíveis a públicos diversos.

No caso específico da Universidade Federal Rural do Rio de Janeiro (UFRRJ), esse quadro se manifesta na forma de iniciativas dispersas de publicação de informações institucionais, frequentemente distribuídas entre diferentes sistemas, páginas e relatórios. A análise de documentos oficiais e do sítio eletrônico da universidade indica a ausência de um portal unificado que reúna, em um único ambiente, um catálogo estruturado de conjuntos de dados institucionais e visualizações analíticas construídas sobre esses dados. Essa fragmentação dificulta tanto o reuso dos dados pela comunidade acadêmica e pela sociedade quanto a construção de uma visão integrada sobre a atuação da instituição.

Este trabalho insere-se nesse contexto ao propor um sistema de portal de dados e transparência para a UFRRJ, concebido como uma solução que integra princípios de dados abertos e práticas de BI. A proposta busca oferecer, em uma plataforma única, um catálogo institucional de conjuntos de dados, alinhado às diretrizes nacionais de dados abertos, e um conjunto de visualizações analíticas que facilitem a compreensão e o reuso dessas informações por diferentes perfis de usuários. Ao fazê-lo, pretende-se contribuir tanto para o fortalecimento da cultura de dados abertos na universidade quanto para a criação de uma base informacional mais robusta para a gestão acadêmica e administrativa.

## **1.1 Objetivo**

O objetivo geral deste trabalho é conceber e desenvolver um portal de dados e transparência para a UFRRJ, integrando a publicação de conjuntos de dados institucionais e visualizações analíticas baseadas em conceitos de BI.

Para alcançar esse objetivo, o estudo analisa o contexto de dados abertos em universidades federais brasileiras, com ênfase em seus PDA e nos modelos adotados para portais institucionais. No caso específico da UFRRJ, busca-se identificar fontes de dados institucionais relevantes e requisitos informacionais associados à transparência e ao apoio à gestão acadêmica e administrativa, de modo a compreender quais bases são prioritárias e como devem ser organizadas.

A partir desse diagnóstico, o trabalho define a arquitetura conceitual e a modelagem de dados de um portal de dados e transparência que integre catálogo de conjuntos de dados e camada analítica baseada em BI. Em seguida, projeta e implementa um protótipo da solução proposta, incluindo mecanismos de ingestão e organização de conjuntos de dados, bem como a integração com uma plataforma de visualização capaz de produzir *dashboards* e painéis interativos.

Por fim, realiza-se um estudo de caso inicial com dados institucionais da UFRRJ, avaliando a viabilidade técnica e institucional do portal e discutindo seu potencial para aprimorar a transparência, a disponibilidade de informações e o uso estratégico de dados na universidade.

## 1.2 Organização do Trabalho

Este trabalho está organizado em cinco capítulos, além desta introdução.

- Capítulo 1: apresenta o contexto da pesquisa, a motivação, o problema abordado, os objetivos geral e específicos e a organização geral do trabalho.
- Capítulo 2: descreve a fundamentação teórica que embasa o estudo. São discutidos conceitos de governo aberto e dados abertos, com foco na Política de Dados Abertos do Poder Executivo federal e na oferta de dados em universidades federais brasileiras. Em seguida, são abordados fundamentos de BI, arquiteturas de dados (como *data warehouses*, *data lakes* e *lakehouses*) e o papel de ferramentas analíticas e de visualização na transformação de dados em informação.
- Capítulo 3: apresenta a proposta do portal de dados e transparência para a

UFRRJ. São detalhadas a motivação específica da solução, a análise de iniciativas relacionadas em instituições congêneres, o desenho do sistema em nível conceitual e a definição dos requisitos, das regras de negócio e dos casos de uso que orientam a sua especificação.

- Capítulo 4: trata do desenvolvimento da solução proposta. São descritas as principais decisões de implementação, a organização dos módulos do sistema e o fluxo de dados entre as camadas de armazenamento, catálogo e visualização, bem como exemplos de *dashboards* e visualizações construídos a partir de dados institucionais.
- Capítulo 5: apresenta as considerações finais do trabalho. São retomados os objetivos propostos, discutidas as contribuições da solução desenvolvida para a transparência e a gestão da informação na UFRRJ, destacadas as limitações identificadas e sugeridas direções para trabalhos futuros que possam aprofundar ou ampliar os resultados alcançados.

# Capítulo 2

## *Análise de dados e Business Intelligence*

Este capítulo apresenta os principais conceitos relacionados à análise de dados e *Business Intelligence* que embasam a proposta deste trabalho. São discutidos os fundamentos de BI, arquiteturas típicas de soluções analíticas, processos de Extração, Transformação e Carga (ETL) e experiências de uso dessas tecnologias em contextos universitários. O objetivo é oferecer ao leitor uma base conceitual suficiente para compreender, nos capítulos seguintes, as decisões de projeto adotadas para o portal institucional de dados e transparência com integração a BI.

### 2.1 Conceitos de *Business Intelligence*

O termo *Business Intelligence* designa um conjunto de processos, metodologias e ferramentas voltados à coleta, integração, análise e disponibilização de informações para apoio à tomada de decisão Loshin (2012). Em geral, soluções de BI buscam transformar grandes volumes de dados operacionais, dispersos em múltiplas fontes, em indicadores e visões consolidadas sobre o desempenho de uma organização.

Loshin (2012) destaca que BI envolve tanto componentes tecnológicos quanto aspectos organizacionais. Além das plataformas de armazenamento e visualização, são

necessários processos de governança de dados, definição de métricas e engajamento de usuários de negócio. Nessa perspectiva, BI não se restringe à geração de relatórios estáticos, mas inclui painéis interativos, *scorecards* e mecanismos de exploração *ad hoc*, que permitem diferentes cortes e agregações sobre os dados, de acordo com as perguntas de negócio formuladas pelos decisores.

Autores como Turban, Sharda e Delen (2014) e Moss e Atre (2003) enfatizam que iniciativas de BI bem-sucedidas dependem da integração entre estratégia organizacional, processos de trabalho e infraestrutura tecnológica. Indicadores mal definidos, dados de baixa qualidade ou falta de envolvimento das áreas usuárias tendem a comprometer o potencial analítico das soluções, mesmo quando a plataforma tecnológica é robusta.

## 2.2 Arquiteturas analíticas e modelagem de dados para BI

Arquiteturas de BI costumam ser organizadas em camadas, envolvendo fontes de dados operacionais, processos de integração e tratamento e, por fim, ferramentas de visualização e análise. Dados provenientes de diferentes sistemas são padronizados e consolidados em um repositório estruturado, a partir do qual podem ser consultados por painéis, relatórios e consultas analíticas (LOSHIN, 2012; CHAUDHURI; DAYAL; NARASAYYA, 2011).

Inmon (2005) descreve o *data warehouse* como um repositório de dados orientado a assuntos, integrado, não volátil e historizado, projetado especificamente para suporte à decisão. A Figura 2.1 ilustra, de forma simplificada, esse tipo de arquitetura, em que dados de múltiplas fontes operacionais, como sistemas transacionais de *ERP*, *CRM*, módulos acadêmicos e financeiros, são extraídos, transformados e carregados por um processo de ETL para um *data warehouse*, que por sua vez alimenta relatórios, consultas *ad hoc* e análises *OLAP*.

Para organizar os dados nesse repositório analítico, Kimball e Ross (2013) propõe a modelagem dimensional, baseada em tabelas de fatos e dimensões. As tabelas de fatos concentram medidas quantitativas, como matrículas, cargas horárias ou

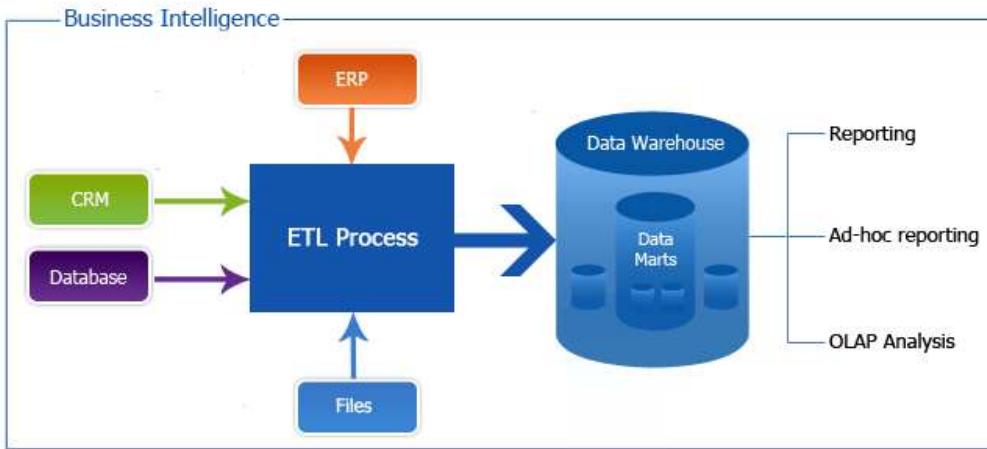


Figura 2.1: Arquitetura simplificada de solução de *Business Intelligence* baseada em processo de ETL e *data warehouse*.

Fonte: Adaptado de CETAX.<sup>1</sup>

valores financeiros, enquanto as dimensões armazenam atributos de contexto, como cursos, unidades acadêmicas, calendários ou categorias temáticas. Essa estrutura favorece a agregação e o *drill-down* de indicadores, além de simplificar a elaboração de consultas por usuários não especialistas em banco de dados.

Embora *data warehouses* sejam tradicionalmente associados a dados estruturados, o crescimento de volumes e formatos de informações levou ao surgimento de arquiteturas mais flexíveis, como *data lakes*. Nesses ambientes, dados estruturados, semiestruturados e não estruturados são armazenados em repositórios distribuídos, com esquemas aplicados apenas no momento da leitura (SAWADOGO; DARMONT, 2021). A Figura 2.2 apresenta uma visão de alto nível de um ecossistema analítico em que um *data lake* centraliza diferentes tipos de dados e serve de base para *data warehouses*, relatórios de BI e aplicações de ciência de dados.

Armbrust et al. (2020) argumentam que abordagens mais recentes buscam combinar as vantagens de *data lakes* e *data warehouses*, dando origem ao conceito de *lakehouse*, que procura conciliar baixo custo de armazenamento em larga escala com mecanismos de governança, desempenho e consistência próximos aos de repositórios analíticos tradicionais. Independentemente da tecnologia adotada, um ponto comum

<sup>1</sup><<https://cetax.com.br/data-warehouse/>>

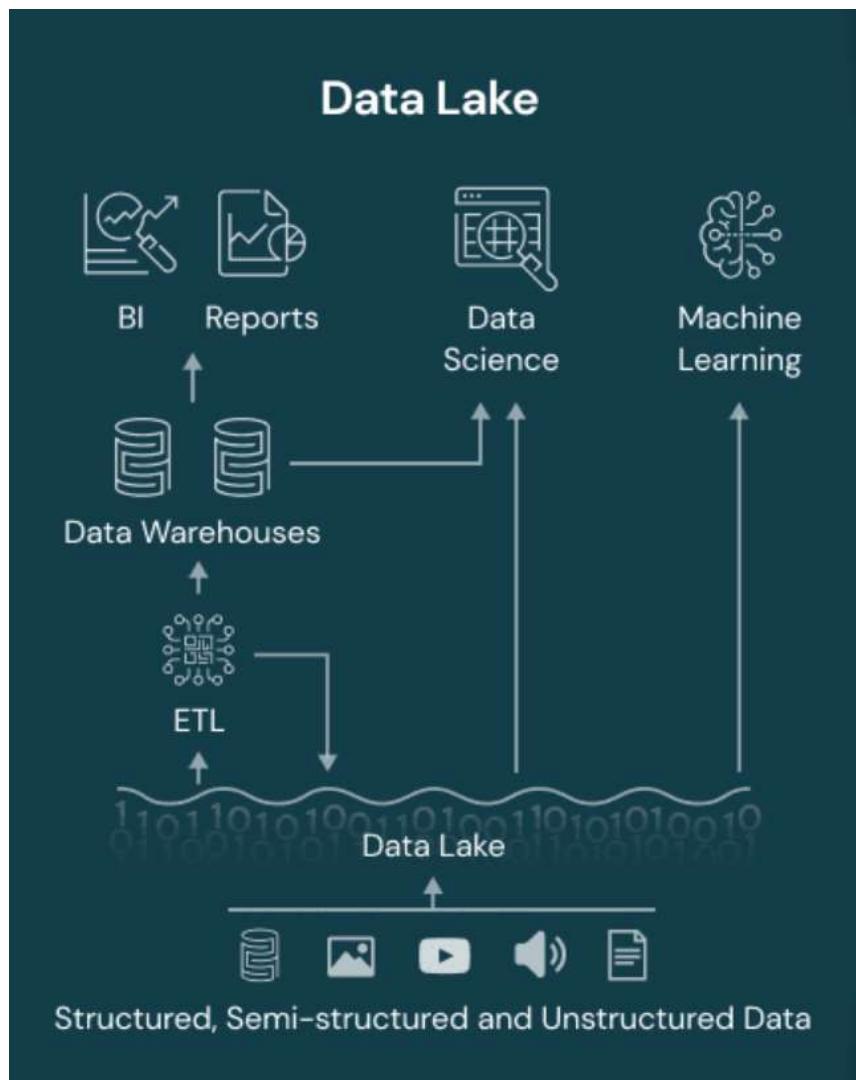


Figura 2.2: Exemplo de ecossistema analítico com *data lake*, *data warehouses* e camadas de análise e BI.

Fonte: Adaptado de Databricks.<sup>2</sup>

nas arquiteturas analíticas é a necessidade de documentar metadados, garantir a qualidade das informações e prover mecanismos de rastreabilidade entre as fontes operacionais e os indicadores publicados.

Em contextos institucionais, como aqueles em que se publicam conjuntos de dados abertos, é comum adotar uma configuração próxima à de um *data warehouse* especializado em *datasets* institucionais. Nesse tipo de arranjo, arquivos tabulares podem ser persistidos em um banco relacional estruturado, com controle de versão,

<sup>2</sup><<https://www.databricks.com/br/blog/2020/01/30/what-is-a-data-lakehouse.html>>

metadados descritivos e registro de operações em log. Plataformas de BI consomem diretamente esse repositório para construir gráficos e painéis interativos, que podem ser incorporados às páginas públicas dos *datasets*. Dessa forma, a arquitetura combina uma camada de ingestão e organização de dados com uma camada analítica, alinhada aos princípios discutidos por Loshin (2012), Kimball e Ross (2013).

## 2.3 Processos de ETL em ambientes de BI

O processo de ETL desempenha papel central na implementação eficaz de soluções de BI, pois garante que os dados extraídos de múltiplas fontes estejam organizados e padronizados antes de serem analisados. Kimball e Ross (2013) descrevem o ETL como um processo composto por três etapas principais, que são extração, transformação e carga.

Na etapa de extração, os dados são coletados de diferentes fontes, como bancos de dados relacionais, sistemas transacionais, serviços expostos por interfaces de programação de aplicações (APIs) e arquivos externos. Essa fase pode apresentar desafios importantes, entre os quais se destacam a integração de dados heterogêneos, a identificação de inconsistências e a manipulação de formatos diversos. Também é fundamental assegurar que os dados estejam atualizados e representem com fidelidade as informações de origem, de modo a preservar a qualidade necessária para as fases subsequentes (INMON, 2005).

A etapa de transformação é responsável pela limpeza, padronização e estruturação dos dados extraídos. Essa fase contribui para a melhoria da qualidade da informação por meio de técnicas como normalização, agregação, tratamento de valores ausentes e padronização de nomenclaturas (MOSS; ATRE, 2003). Também podem ser realizadas conversões de tipos de dados, validações de integridade e enriquecimento com atributos derivados. Esses procedimentos buscam garantir uniformidade e consistência para os dados que serão armazenados no repositório analítico.

A etapa de carga consiste na inserção dos dados transformados em um repositório final, geralmente um *data warehouse* ou um *data lake*. Segundo Chaudhuri, Dayal

e Narasayya (2011), essa etapa pode ocorrer de forma incremental, quando apenas novos registros ou alterações são aplicados, ou por meio de cargas completas, em que todos os dados são substituídos periodicamente. A escolha entre essas estratégias depende de fatores como volume de dados, requisitos de atualização e restrições de janelas de processamento.

A Figura 2.3 apresenta, de maneira esquemática, o ciclo de ETL com suas três fases básicas e o fluxo de dados das fontes até o repositório de destino.

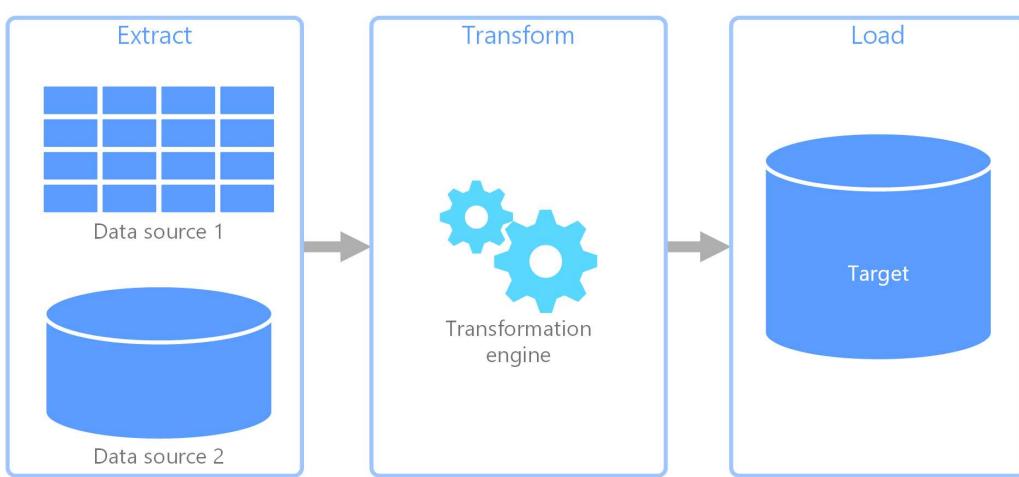


Figura 2.3: Visão esquemática do processo de ETL.

Fonte: Adaptado de Microsoft Learn.<sup>3</sup>

O sucesso de um sistema de BI depende diretamente da qualidade dos dados analisados, o que torna o ETL um componente crítico desse processo. Kimball e Ross (2013) e Loshin (2012) observam que dados inconsistentes ou mal estruturados podem comprometer a confiabilidade das análises e levar a decisões equivocadas. Processos de ETL bem definidos contribuem para eliminar duplicidades, preencher lacunas, padronizar formatos e consolidar informações provenientes de diferentes sistemas em uma visão unificada.

<sup>3</sup><<https://learn.microsoft.com/pt-br/azure/architecture/data-guide/relational-data/etl>>

## 2.4 *Business Intelligence* na gestão universitária

Diversos trabalhos apontam o potencial de BI para apoiar a gestão acadêmica e administrativa em instituições de ensino superior. Júnior et al. (2022) relatam a implantação de painéis de indicadores em uma universidade federal, integrando dados de módulos acadêmicos e administrativos para monitorar matrículas, evasão, carga horária docente e outros aspectos relevantes para o planejamento institucional. Os autores destacam benefícios como maior visibilidade sobre o desempenho da instituição, facilitação da comunicação entre setores e suporte a decisões baseadas em evidências.

De forma semelhante, Vieira, Seabra e Garcés (2025) descrevem o uso de BI em sistemas de gestão acadêmica, com enfoque na construção de *dashboards* que permitem acompanhar trajetórias estudantis, resultados em disciplinas, fluxos de permanência e conclusão de cursos. Entre os resultados observados estão a identificação mais rápida de situações de risco de evasão, o acompanhamento de metas institucionais e a ampliação da transparência interna sobre indicadores-chave.

Essas experiências evidenciam que soluções de BI podem ser aplicadas tanto a dados operacionais internos quanto a informações que interessam ao público externo, como estatísticas de vagas, matrículas e concluintes, produção acadêmica, execução orçamentária e indicadores de assistência estudantil. Ao integrar essas capacidades analíticas a um portal institucional de dados, abre-se a possibilidade de combinar transparência pública com mecanismos de exploração visual que facilitam a interpretação das informações por diferentes perfis de usuários.

## 2.5 Síntese da fundamentação teórica

A fundamentação apresentada neste capítulo buscou discutir conceitos centrais de análise de dados e *Business Intelligence* relevantes para o desenvolvimento do portal de dados e transparência proposto. Inicialmente, foram definidos BI e seus componentes principais, ressaltando a necessidade de integração entre tecnologia,

processos e áreas de negócio (LOSHIN, 2012; TURBAN; SHARDA; DELEN, 2014). Em seguida, foram exploradas arquiteturas analíticas baseadas em *data warehouses*, modelagem dimensional e, de forma complementar, abordagens recentes envolvendo *data lakes* e *lakehouses* (INMON, 2005; KIMBALL; ROSS, 2013; SAWADOGO; DARMONT, 2021; ARMBRUST et al., 2020).

Também foram detalhados os processos de ETL, com ênfase em suas etapas de extração, transformação e carga e em sua importância para a garantia de qualidade dos dados utilizados em soluções de BI (INMON, 2005; MOSS; ATRE, 2003; CHAUDHURI; DAYAL; NARASAYYA, 2011; KIMBALL; ROSS, 2013; LOSHIN, 2012). Por fim, foram discutidas aplicações de BI na gestão universitária, destacando experiências de implantação de painéis de indicadores em universidades federais brasileiras (JÚNIOR et al., 2022; VIEIRA; SEABRA; GARCÉS, 2025).

Essa base teórica fundamenta as decisões de projeto detalhadas no Capítulo 3, em que se descrevem os requisitos, a modelagem de dados e os mecanismos de integração entre o portal de transparência e a plataforma de BI, concebidos como uma solução analítica voltada tanto ao apoio à gestão interna quanto ao fortalecimento da transparência pública.

# Capítulo 3

## Proposta

A transparéncia de dados constitui eixo estruturante da gestão pública e do compromisso social das instituições de ensino superior. Um portal de dados e transparéncia permite reduzir assimetrias de informação, ampliar o controle social e fortalecer a prestação de contas, ao mesmo tempo em que favorece o reuso por pesquisadores, jornalistas e pela comunidade. Contudo, a produção informacional costuma estar dispersa entre setores, em formatos heterogêneos e com baixa padronização, o que dificulta a consolidação, a interpretação estratégica e a atualização periódica do conteúdo disponibilizado ao público.

No contexto universitário, esses desafios se traduzem em múltiplas bases, sistemas departamentais e planilhas paralelas que tratam de temas como ensino, pesquisa, extensão, orçamento e gestão de pessoas. Muitas vezes, informações importantes aparecem fragmentadas em relatórios estáticos ou em documentos produzidos sob demanda, sem um catálogo institucional estruturado que permita localizar, comparar e acompanhar indicadores ao longo do tempo. Mesmo quando há seções de transparéncia ou iniciativas pontuais de dados abertos, falta um mecanismo que articule de forma consistente a produção interna de dados, sua publicação em formato aberto e a disponibilização de visualizações analíticas acessíveis a diferentes perfis de usuário.

Ao mesmo tempo, abordagens de *Business Intelligence* (BI) têm ganhado espaço como forma de integrar dados dispersos, construir painéis de indicadores e apoiar

gestores na análise e no acompanhamento de resultados institucionais. Na prática, entretanto, é comum que essas soluções permaneçam restritas a ambientes internos, com acesso limitado a usuários especializados e pouca integração com os portais públicos de transparência. Painéis de BI e catálogos de dados abertos acabam sendo mantidos em espaços distintos, dificultando que o cidadão tenha, em um único lugar, acesso simultâneo aos dados brutos, ao contexto informacional e às visualizações analíticas associadas.

Diante desse contexto, este capítulo transita do diagnóstico para a construção de uma solução, apresentando a proposta deste trabalho: o desenvolvimento de um sistema institucional de publicação e análise de dados orientado por princípios de governança da informação e de BI. A proposta busca suprir a ausência, ainda recorrente em muitas universidades, de um mecanismo unificado para organizar, catalogar e comunicar informações institucionais com confiabilidade e clareza, contribuindo para maior transparência, efetividade na tomada de decisão e aproximação com a sociedade.

A proposta contempla a definição de requisitos funcionais e não funcionais e de regras de negócio que asseguram coerência aos processos de ingestão, catalogação, publicação, atualização e descontinuação de conjuntos de dados. Descrevem-se os principais casos de uso associados a esses processos e apresenta-se uma visão conceitual da solução, articulando seus componentes e a forma como se integra a ferramentas de análise e visualização. Antes da especificação, são discutidos trabalhos relacionados que situam o estado da arte em termos de portais institucionais de dados abertos e transparência no âmbito das instituições de ensino superior.

### **3.1 Motivação**

A motivação para a proposta desenvolvida neste trabalho está associada tanto a exigências gerais de transparência e dados abertos no setor público quanto a necessidades específicas das instituições de ensino superior. No plano mais amplo, o Decreto nº 8.777, de 11 de maio de 2016, institui a Política de Dados Abertos do

Poder Executivo federal explicita objetivos como promover a publicação de bases em formato aberto, aprimorar a cultura de transparência e franquear aos cidadãos o acesso, de forma aberta, aos dados produzidos ou acumulados pelos órgãos públicos (BRASIL, 2016). Entre os instrumentos dessa política destacam-se os Planos de Dados Abertos e o Portal Brasileiro de Dados Abertos, que organizam informações em catálogos de conjuntos de dados acompanhados de metadados e recursos em formatos processáveis por máquina, consolidando um modelo de referência para a publicação estruturada de dados governamentais (BRASIL, 2025).

No âmbito da Universidade Federal Rural do Rio de Janeiro (UFRRJ), a página institucional sobre a Lei de Acesso à Informação explicita o compromisso da universidade com a transparência ativa, a divulgação de informações de interesse coletivo e a disponibilização de dados em formatos acessíveis ao cidadão. Esse compromisso se desdobra na elaboração de um Plano de Dados Abertos, que define conjuntos de dados prioritários, responsabilidades internas e cronogramas de implementação (UFRRJ, 2020), bem como em relatórios de monitoramento que registram avanços e dificuldades na execução das ações previstas (UFRRJ, 2023). Esses documentos evidenciam tanto a importância institucional da agenda de dados abertos quanto desafios práticos relacionados à identificação de bases relevantes, à padronização de informações e à atualização periódica dos conteúdos publicados.

Embora instituições de ensino superior possuam arranjos administrativos próprios, Albano, Craveiro e Gaffrée (2021) e Pires (2019) apontam que elas são pressionadas por expectativas semelhantes de transparência, participação social e uso de evidências na gestão. Estudos sobre oferta de dados abertos em universidades federais indicam que a implementação da política ainda é heterogênea, com diferenças significativas na quantidade de conjuntos de dados disponibilizados, na qualidade dos metadados e na regularidade de atualização das bases. Em muitos casos, informações relevantes aparecem diluídas em relatórios estáticos, documentos normativos e planilhas pouco padronizadas, o que dificulta o reúso e a análise sistemática por parte da sociedade e da própria gestão universitária, como discutido por (ALBANO; CRAVEIRO; GAFFRÉE, 2021) e (LIMA; PIRES, 2022).

Paralelamente, abordagens de BI, discutidas no Capítulo 2, vêm sendo exploradas na educação superior sobretudo para consolidar indicadores acadêmicos e administrativos a partir de dados dispersos. Estudos como Júnior et al. (2022) e Vieira, Seabra e Garcés (2025) mostram que painéis de BI têm sido utilizados para monitorar matrículas, evasão, carga horária docente e trajetórias estudantis, a partir da integração de informações provenientes de sistemas legados e bases administrativas. Em síntese, essas experiências indicam que universidades já dispõem, em muitos casos, de infraestruturas analíticas internas baseadas em BI (LOSHIN, 2012), ainda que nem sempre articuladas de forma sistemática com os portais de transparência e com a publicação de dados abertos.

Apesar desses avanços, diversos estudos apontam que grande parte dos portais de transparência e dados abertos ainda privilegia a oferta de arquivos para *download*, com apoio limitado a visualizações interativas e a mecanismos que integrem de forma explícita dados brutos, metadados e painéis de BI. Nesse cenário, como observam Albano, Craveiro e Gaffrée (2021) e Lima e Pires (2022), a produção de *dashboards* tende a ficar restrita a equipes específicas, com processos de extração e consolidação realizados de forma manual, o que dificulta a reprodução de análises, aumenta o risco de inconsistências e reduz a transparência metodológica dos resultados apresentados. Além disso, as ferramentas de BI são frequentemente utilizadas em ambientes separados dos portais institucionais, com acesso restrito e interfaces pouco adaptadas ao público em geral, de modo que a integração web entre BI e transparência costuma ser limitada ou inexistente.

Nesse contexto, ganha relevância a concepção de um sistema que integre, em uma solução única, três dimensões complementares. A primeira é a organização da informação em conjuntos de dados institucionais, associados a metadados claros, unidades responsáveis e ciclos de atualização definidos, em consonância com princípios de dados abertos e governança da informação. A segunda é a disponibilização desses conjuntos de dados em formatos abertos, com mecanismos de busca, filtragem e *download* que facilitem o reúso por diferentes públicos, internos e externos. A terceira é a articulação com ferramentas de BI, de forma que visualizações analíticas e painéis interativos possam ser diretamente associados aos conjuntos de dados

e disponibilizados no próprio portal de transparéncia, aproximando dados brutos, contexto informacional e leitura analítica em um mesmo ambiente. A proposta deste trabalho assume, assim, a necessidade de superar a separação tradicional entre portais de dados abertos e ambientes internos de BI, desenhando um sistema em que a camada analítica possa ser incorporada, de maneira integrada, à interface pública de transparéncia.

A proposta apresentada neste capítulo busca responder a essa convergência. Ao tomar como referência o marco de dados abertos e as experiências existentes de uso de BI em instituições de ensino, mas ao mesmo tempo reconhecer as limitações de portais que se restringem à publicação de arquivos isolados e de soluções de BI pouco integradas à web, o trabalho propõe um sistema institucional de publicação e análise de dados orientado à realidade das instituições de ensino superior. A expectativa é que uma solução que combine catálogo de conjuntos de dados, controle de versões, disponibilização em formato aberto e integração nativa com painéis de BI possa contribuir não apenas para atender exigências de transparéncia, mas também para fortalecer uma cultura de gestão baseada em dados, capaz de apoiar decisões acadêmicas e administrativas de forma mais sistemática e verificável.

## **3.2 Trabalhos relacionados**

A adoção de políticas de dados abertos em universidades públicas brasileiras se materializa, em grande medida, na criação de portais institucionais que organizam e divulgam informações em formato reutilizável. Esses portais costumam articular dois instrumentos principais, o plano de dados abertos e a plataforma de publicação. O plano define prioridades de abertura, cronogramas e responsabilidades internas, enquanto o portal funciona como catálogo de conjuntos de dados acompanhados de metadados e arquivos para *download*, como descrito nos documentos de planejamento de dados abertos de universidades federais (UFG, 2023; UFU, 2023).

Um exemplo representativo é o Portal de Dados Abertos da Universidade Federal de Goiás, que se apresenta como ferramenta para tornar dados institucionais acessíveis

e utilizáveis, com ênfase em simplicidade e organização para que qualquer pessoa encontre com facilidade as informações de que precisa. O portal declara o objetivo de disponibilizar diferentes tipos de dados no âmbito da universidade, incluindo ensino, pesquisa, extensão, materiais, patrimônio e processos, funcionando como um catálogo que centraliza conjuntos de dados publicados pelas pró-reitorias (UFG, 2023). O acesso público a esse catálogo é realizado por meio de interface web específica para dados abertos.<sup>1</sup>

A Universidade Federal do Cariri oferece um portal de dados abertos sustentado por um plano institucional que busca promover transparência e comunicação com a comunidade acadêmica e a sociedade (UFCA, 2025a). O portal é apresentado como plataforma dedicada ao acesso à informação pública e utiliza *software* de catálogo de dados com suporte ao conceito de *dataset*, indicando para cada conjunto de dados metadados, arquivos associados e possibilidades de acesso automatizado (UFCA, 2025b).<sup>2</sup> Outras universidades federais, como a Universidade Federal do Ceará e a Universidade Federal do Acre, divulgam planos de dados abertos que também reforçam a importância de catálogos institucionais e de critérios para seleção de bases prioritárias, ainda que os portais estejam em diferentes estágios de implementação (UFC, 2024; UFAC, 2024).

A Universidade Federal de Uberlândia mantém portal específico de dados abertos integrado ao ecossistema *gov.br*, no qual os dados são organizados em conjuntos associados a uma organização e a categorias temáticas UFU (2025). Em documentos institucionais, o plano de dados abertos é destacado como instrumento para dar visibilidade e transparência às ações da universidade, com ênfase em bases relacionadas a servidores, estudantes e cursos, que passam a ser disponibilizadas publicamente na internet (UFU, 2023). Também nesse caso, o conceito de conjunto de dados aparece como unidade de publicação, combinando descrição, órgão responsável e arquivos em formatos como CSV, conforme ilustrado na Figura 3.1.<sup>3</sup>

Em comum, essas iniciativas adotam a ideia de catálogo de conjuntos de dados,

---

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

<sup>2</sup><<https://dados.ufca.edu.br/>>

<sup>3</sup><<https://dados.ufu.br/>>

Figura 3.1: Exemplo de catálogo de conjuntos de dados em portal institucional de dados abertos.

Fonte: UFU (2025).

em que cada *dataset* é descrito por um conjunto mínimo de metadados, como título, descrição, unidade responsável, período de referência e formatos disponíveis. A ênfase recai na disponibilização de arquivos para *download*, frequentemente em CSV ou JSON, e na conformidade com diretrizes de transparência ativa, como indicam (UFG, 2025; UFU, 2025). Alguns portais começam a incorporar visualizações básicas, como gráficos e painéis simples, mas em muitos casos a camada analítica permanece separada ou limitada a relatórios estáticos, o que reforça a distância entre o catálogo de dados e as ferramentas de BI utilizadas em ambientes internos.

O sistema proposto neste trabalho se aproxima dessas experiências ao adotar o conceito de conjunto de dados como unidade central de publicação e ao organizar um catálogo de informações acompanhado de metadados institucionais. Ao mesmo tempo, distingue-se por enfatizar mecanismos de ingestão estruturada de arquivos tabulares, controle explícito de versões e associação sistemática entre *datasets* e visualizações analíticas produzidas em plataforma de BI. Essa combinação busca aproximar, em uma única solução, as práticas de transparência e de dados abertos já consolidadas em universidades públicas e a exploração analítica por meio de painéis interativos, reduzindo a distância entre dados brutos, metadados descritivos

e instrumentos de apoio à tomada de decisão. Diferentemente de arranjos em que a BI permanece restrita a ambientes internos, o desenho proposto assume, desde a concepção, que as visualizações analíticas devem ser integradas ao próprio portal de dados, de forma que a experiência do usuário não seja fragmentada entre múltiplos sistemas.

### **3.3 Especificação da solução**

Esta seção descreve os elementos que orientam a especificação do sistema de publicação e análise de dados institucionais proposto. São apresentados, inicialmente, a visão geral do desenho do sistema e os atores envolvidos. Em seguida, são descritos os requisitos funcionais e não funcionais, as regras de negócio e os principais casos de uso da aplicação. Em conjunto, esses elementos definem o comportamento esperado do sistema e servem de base para as etapas de modelagem e implementação.

#### **3.3.1 Visão geral do desenho do sistema**

Do ponto de vista conceitual, o sistema proposto é organizado em três camadas principais. A primeira é a camada de gestão interna de dados, que reúne as funcionalidades utilizadas por docentes, servidores técnico-administrativos e analistas para cadastrar conjuntos de dados institucionais, carregar arquivos em formato tabular, manter o histórico de versões e registrar metadados essenciais, como unidade responsável, período de referência e categoria temática. Nessa camada concentram-se as atividades de preparação, validação e atualização dos dados que serão disponibilizados ao público.

A segunda camada é a de serviços de dados, responsável por estruturar e disponibilizar os conjuntos de dados consolidados de forma organizada, garantindo que possam ser reutilizados por diferentes aplicações analíticas. Essa camada preserva metadados, histórico de versões e regras de atualização, além de expor os *datasets* de modo padronizado para que possam ser consumidos por ferramentas de análise e visualização.

A terceira camada é a de publicação e análise, que corresponde ao portal de transparência acessado pelo público. Nessa camada, os conjuntos de dados são apresentados em forma de catálogo, com mecanismos de busca e filtragem, descrição detalhada e opções de *download* em formato aberto. A mesma interface reúne, para cada *dataset*, as visualizações analíticas associadas, permitindo que o usuário público explore gráficos e painéis interativos diretamente no portal.

A interação entre essas camadas acompanha o ciclo de vida dos dados institucionais. Na gestão interna, docentes e analistas organizam e atualizam os conjuntos de dados. Na camada de serviços, essas informações são estruturadas e expostas de forma consistente. Por fim, na camada de publicação e análise, o cidadão acessa o catálogo, consulta metadados, realiza *downloads* e interage com painéis analíticos. A integração com a ferramenta de BI é concebida, em nível conceitual, como um mecanismo em que os conjuntos de dados estruturados pelo portal servem de base para as visualizações, que por sua vez são associadas de forma sistemática às páginas públicas dos *datasets*. Dessa forma, o desenho do sistema busca articular, em uma única solução, os processos de ingestão, organização, publicação e exploração analítica dos dados, respondendo às limitações observadas em arranjos em que a BI permanece separada dos portais de transparência e em que as próprias plataformas analíticas não oferecem, isoladamente, meios adequados de publicação voltados ao público externo.

Em termos arquiteturais, esse arranjo configura uma arquitetura em camadas sobre o modelo cliente–servidor, na qual a interface pública atua como cliente, a camada de serviços de dados concentra regras de negócio e controle de acesso às informações institucionais, e a camada de gestão interna dá suporte aos fluxos administrativos de cadastro e atualização dos conjuntos de dados. Essa organização favorece a separação de responsabilidades e contribui para requisitos de segurança, manutenibilidade e evolução futura das funcionalidades do portal, em consonância com os requisitos funcionais, não funcionais e regras de negócio definidos nas subseções seguintes.

### 3.3.2 Decisões arquiteturais da proposta

A especificação do sistema proposto envolve algumas decisões arquiteturais que condicionam a forma como os requisitos serão atendidos. Essas decisões podem ser registradas e detalhadas em artefatos específicos de documentação, de modo a preservar o histórico de alternativas consideradas e os critérios que orientaram as escolhas. Entre as decisões mais relevantes, destacam-se:

- **Centralização no conceito de conjunto de dados institucional:** o sistema adota o *dataset* como unidade central de publicação, descrita por metadados institucionais e associada a versões de arquivos tabulares. Essa decisão está alinhada às diretrizes de dados abertos e facilita a rastreabilidade entre as informações divulgadas e suas fontes administrativas, respondendo aos requisitos de organização, governança da informação e integridade dos dados.
- **Separação entre gestão interna, serviços de dados e publicação pública:** a solução organiza-se em três camadas conceituais, reservando uma camada para os fluxos administrativos de cadastro e atualização de dados, outra para a consolidação e exposição estruturada dos *datasets* e uma terceira para o acesso público. Essa organização favorece a separação de responsabilidades e contribui diretamente para os requisitos de segurança, manutenibilidade, escalabilidade e controle de acesso definidos para o portal.
- **Integração entre catálogo de dados e camada analítica:** em vez de manter painéis de BI em ambiente totalmente separado, a proposta prevê a associação sistemática entre *datasets* e visualizações analíticas, de forma que os painéis sejam exibidos diretamente nas páginas de cada conjunto de dados. Essa decisão busca atender simultaneamente aos requisitos de transparência, reproduzibilidade das análises, usabilidade para diferentes perfis de usuário e aproveitamento das capacidades das plataformas analíticas utilizadas pela instituição.
- **Controle explícito de versões:** o sistema registra cada atualização de arquivo como uma nova versão do *dataset*, preservando o histórico e evitando a

perda de informações anteriores. Essa decisão está diretamente relacionada aos requisitos de integridade, auditoria, rastreabilidade e preservação de histórico, ao mesmo tempo em que apoia práticas de governança de dados voltadas à transparência sobre alterações realizadas ao longo do tempo.

- **Separação entre responsabilidades administrativas e publicação ao público:** a definição de papéis distintos para responsáveis pelo cadastro de dados e para os agentes autorizados a publicar, despublicar ou arquivar *datasets* reflete uma opção arquitetural por fluxos de validação e revisão antes da exposição pública das informações. Essa decisão contribui para os requisitos de segurança, conformidade legal e qualidade da informação, reduzindo o risco de divulgação indevida de dados sensíveis ou incorretos.

Essas decisões estruturam a forma como o sistema será detalhado nas camadas de *software* e nos componentes tecnológicos descritos em capítulos posteriores, e se relacionam diretamente aos requisitos funcionais, não funcionais e regras de negócio apresentados a seguir.

### 3.3.3 Atores e perfis de uso

O sistema envolve diferentes perfis de uso, que participam de etapas complementares do ciclo de vida dos dados. A seguir, são descritos os principais atores considerados na especificação da solução.

- a) **Docente ou servidor técnico-administrativo responsável:** prepara arquivos em formato tabular com dados institucionais sob sua responsabilidade, como informações de ensino, pesquisa ou extensão. Cabe a esse ator realizar o *upload* desses arquivos no portal, preencher os metadados mínimos do conjunto de dados e solicitar a publicação ou atualização das informações, quando autorizado.
- b) **Analista de BI:** profissional com perfil técnico em análise e visualização de dados, encarregado de construir consultas, gráficos e painéis em uma plataforma de BI, a partir dos conjuntos de dados disponibilizados pelo portal. Também cabe a esse ator registrar no sistema as visualizações associadas a cada *dataset* e acompanhar o

ciclo de vida desses painéis.

c) **Usuário público:** qualquer cidadão, membro da comunidade acadêmica ou parte interessada que acessa o portal de transparência para localizar conjuntos de dados, interagir com painéis analíticos e realizar *download* de arquivos em formato aberto.

d) **Plataforma de BI:** sistema externo de análise e visualização de dados que consome os conjuntos de dados estruturados pelo portal e fornece visualizações como gráficos, tabelas e *dashboards*. Esse ator participa principalmente dos casos de uso relacionados à associação e exibição de visualizações analíticas, atuando em conjunto com o analista de BI para que o usuário público interaja com os painéis.

#### **3.3.4 Requisitos funcionais**

Os Requisitos Funcionais (RF) definem as funcionalidades que o sistema deverá disponibilizar aos usuários, especificando o comportamento esperado diante de diferentes ações. Esses requisitos buscam garantir que o sistema atenda às necessidades operacionais e institucionais, orientando o processo de desenvolvimento e validação.

Os principais requisitos funcionais identificados para o portal de transparência são:

a) (RF01) Autenticação de usuários administrativos: o sistema deve permitir a autenticação de usuários com perfil administrativo, como docentes ou analistas de BI designados, de modo a restringir o acesso às funcionalidades de gestão.

b) (RF02) Gestão de usuários e permissões: o sistema deve possibilitar o cadastro, edição e desativação de usuários internos, bem como a associação a perfis de acesso adequados às responsabilidades assumidas no portal.

c) (RF03) Cadastro de conjuntos de dados (*datasets*): o sistema deve permitir o cadastro de conjuntos de dados institucionais, contemplando metadados como título, descrição, unidade responsável, categoria temática, período de referência, palavras-chave e status.

- d) (RF04) *Upload* de arquivos tabulares: o sistema deve permitir o *upload* de arquivos em formato tabular, como CSV, associados a um *dataset*, realizando validações de estrutura, tais como cabeçalho, delimitador, codificação e tipos elementares.
- e) (RF05) Armazenamento estruturado dos dados: o sistema deve armazenar os dados contidos nos arquivos carregados em repositório estruturado, mantendo a associação com o *dataset* e com a versão correspondente.
- f) (RF06) Controle de versão de arquivos: o sistema deve manter histórico de versões de um mesmo *dataset*, registrando sucessivos *uploads* sem perda das versões anteriores e indicando datas de criação e atualização.
- g) (RF07) Publicação, despublicação e arquivamento: o sistema deve permitir alterar o status dos *datasets* entre rascunho, publicado e arquivado, controlando a visibilidade das informações na interface pública.
- h) (RF08) Disponibilização de dados para *download*: para *datasets* publicados, o sistema deve disponibilizar os dados em formato aberto, permitindo o *download* direto pelo usuário público.
- i) (RF09) Exposição dos dados a uma plataforma de BI: o sistema deve tornar os *datasets* acessíveis a uma plataforma externa de análise e visualização de dados, permitindo sua utilização em consultas, gráficos e painéis.
- j) (RF10) Cadastro de visualizações de BI: o sistema deve permitir o registro, junto a cada *dataset*, de visualizações provenientes da plataforma de BI, como *cards* e *dashboards*, armazenando identificadores e descrições necessárias à sua exibição.
- k) (RF11) Exibição de painéis na interface pública: o portal de transparência deve exibir, nas páginas dos *datasets* publicados, os painéis interativos associados, de acordo com os recursos oferecidos pela plataforma de BI.
- l) (RF12) Busca e filtragem de *datasets*: o sistema deve oferecer mecanismos de busca e filtragem de *datasets* por palavra-chave, categoria temática, unidade responsável, período de referência e outros critérios relevantes.
- m) (RF13) Registro de atividades (auditoria): o sistema deve registrar ações

administrativas importantes, como criação, edição, publicação, *upload* e arquivamento de dados, associando as operações a usuário, data e hora.

### **3.3.5 Requisitos não funcionais**

Além das funcionalidades específicas que o sistema deve oferecer, é essencial considerar os Requisitos Não Funcionais (RNF), que definem atributos de qualidade, desempenho, segurança e usabilidade. Esses requisitos não descrevem diretamente o que o sistema faz, mas como ele deve se comportar e quais características deve possuir para garantir confiabilidade, eficiência e boa experiência de uso.

Os principais requisitos não funcionais considerados para o portal são:

- a) (RNF01) Segurança e controle de acesso: em complemento aos requisitos funcionais de autenticação e autorização (RF01 e RF02), os acessos às funcionalidades administrativas devem ser protegidos por mecanismos seguros de autenticação e autorização, de modo a evitar alterações não autorizadas em dados e configurações.
- b) (RNF02) Confiabilidade e integridade dos dados: o sistema deve garantir que os dados não sejam corrompidos durante os processos de *upload*, armazenamento e leitura, mantendo consistência entre versões e registros de auditoria.
- c) (RNF03) Desempenho: as páginas públicas do portal devem ser carregadas em poucos segundos, de forma que, em condições usuais de acesso por banda larga, pelo menos noventa e cinco por cento das requisições apresentem tempo de resposta de até três segundos.
- d) (RNF04) Usabilidade e acessibilidade: a interface deve adotar princípios de simplicidade, consistência e linguagem clara, alinhada a diretrizes de acessibilidade digital, facilitando o uso por públicos com diferentes níveis de letramento digital.
- e) (RNF05) Interoperabilidade: os dados devem ser disponibilizados em formatos abertos e padronizados, permitindo o reuso em ferramentas externas de análise, bem como a futura disponibilidade por meio de interfaces de acesso automatizado por outros sistemas.

f) (RNF06) Escalabilidade e manutenibilidade: o sistema deve permitir a inclusão de novos *datasets* e a evolução de funcionalidades sem necessidade de alterações extensas nas partes já existentes, favorecendo uma organização modular dos componentes de software e a manutenção ao longo do tempo.

g) (RNF07) Conformidade legal: o sistema deve observar a legislação vigente relacionada à transparência pública e à proteção de dados pessoais, garantindo que apenas informações adequadas à divulgação sejam disponibilizadas.

#### **3.3.6 Regras de negócio**

As Regras de Negócio (RN) deste sistema de publicação e análise de dados traduzem as diretrizes e restrições inerentes ao domínio analisado, garantindo a integridade das informações e a coerência dos fluxos operacionais. Ao definir critérios claros de validação e controle, essas regras preservam a qualidade da informação que será manipulada e apresentada pelo portal.

As principais regras de negócio identificadas são:

a) (RN01) Metadados obrigatórios: todo *dataset* deve conter, no mínimo, título, descrição, unidade responsável, categoria temática e período de referência para ser registrado no sistema.

b) (RN02) Condição para publicação: um *dataset* só pode ser publicado quando houver pelo menos um arquivo tabular validado e todos os metadados obrigatórios estiverem completos.

c) (RN03) Papéis e responsabilidades: a publicação, despublicação e o arquivamento de *datasets* devem ser realizados por usuários com perfil administrativo, como analistas de BI ou docentes e servidores formalmente designados para essa função. Demais responsáveis podem cadastrar e editar *datasets* e realizar *uploads*, permanecendo sob responsabilidade do perfil administrativo a liberação para o público.

d) (RN04) Tratamento de dados sensíveis: dados pessoais ou sensíveis não devem

ser carregados no sistema. O responsável pelo *dataset* deve realizar previamente a anonimização, agregação ou supressão dessas informações, de acordo com a legislação vigente.

- e) (RN05) Preservação de histórico: a exclusão física de *datasets* e versões deve ser evitada. Quando um conjunto de dados deixar de ser adequado à divulgação, deve ser marcado como arquivado, mantendo o registro para fins de histórico e auditoria.
- f) (RN06) Unicidade de identificação: cada *dataset* deve possuir identificador único, evitando ambiguidade quanto à sua origem e conteúdo, mesmo quando pertence à mesma categoria temática de outros conjuntos de dados.
- g) (RN07) Periodicidade de atualização: sempre que pertinente, cada *dataset* deve registrar a periodicidade mínima de atualização, como anual, semestral ou mensal, permitindo o acompanhamento da atualidade da informação e a detecção de conteúdos potencialmente desatualizados.
- h) (RN08) Compatibilidade estrutural entre versões: alterações significativas na estrutura de colunas de um *dataset* devem ser explicitamente confirmadas pelo usuário responsável no momento do *upload*, sendo registradas como mudança de versão e sinalizadas no histórico.
- i) (RN09) Associação entre *datasets* e visualizações: apenas *datasets* publicados podem ter visualizações de BI exibidas no portal público. A despublicação de um *dataset* implica a retirada de seus painéis associados da interface destinada ao usuário público.

### **3.3.7 Casos de uso e fluxos principais**

A seguir, são apresentados os principais casos de uso do sistema de publicação e análise de dados, considerando os atores previamente identificados: docente ou servidor responsável, analista de BI, usuário público, plataforma de BI e o próprio sistema. Esses casos de uso descrevem, em nível conceitual, como o sistema deve se comportar em situações típicas de cadastro, publicação, visualização e consulta de informações.

**a) UC01: Cadastrar conjunto de dados**

**Autor principal:** Docente ou servidor responsável.

**Pré-condição:** o autor principal deve estar autenticado no sistema.

**Referências cruzadas:** RF03, RF04, RF05, RF06, RF13, RNF01, RNF02, RN01, RN02, RN04, RN06, RN07, RN08.

**Fluxo principal:**

1. O responsável acessa a área administrativa do portal.
2. O responsável seleciona a opção de criar um novo conjunto de dados.
3. O sistema apresenta o formulário de cadastro de metadados do *dataset*.
4. O responsável informa os metadados obrigatórios, como título, descrição, unidade responsável, categoria temática e período de referência e, quando necessário, metadados adicionais.
5. O responsável seleciona o arquivo tabular a ser carregado para o *dataset*.
6. O sistema realiza a validação estrutural do arquivo, verificando cabeçalho, delimitador, codificação e tipos elementares, e apresenta ao responsável um resumo dos campos e tipos identificados, cabendo a ele confirmar que os tipos de dados foram reconhecidos corretamente antes de prosseguir.
7. O sistema armazena os dados do arquivo em repositório estruturado, associando os registros ao *dataset* e registrando uma nova versão.
8. O sistema registra o *dataset* em estado de rascunho e grava a operação no registro de auditoria.
9. O sistema apresenta um resumo do cadastro ao responsável.
10. O caso de uso é encerrado.

**Fluxo de exceção: metadados obrigatórios incompletos**

1. Se o responsável tentar prosseguir sem preencher todos os metadados obrigatórios, o sistema informa que existem campos pendentes, registra a tentativa no log de auditoria e destaca os campos que precisam ser corrigidos.
2. O sistema retorna ao formulário de cadastro para que o responsável conclua o preenchimento.
3. O fluxo retorna ao passo 4 do fluxo principal.

**Fluxo de exceção: arquivo tabular inválido**

1. Se o arquivo não atender às validações estruturais, o sistema apresenta

mensagem com o erro identificado e regista a ocorrência no log de auditoria. 2. O sistema não regista a versão do arquivo. 3. O fluxo retorna ao passo 5 do fluxo principal para que o responsável selecione um novo arquivo.

**Pós-condição:** existe um novo *dataset* registrado no sistema, em estado de rascunho, com metadados obrigatórios preenchidos e pelo menos uma versão de arquivo associada.

#### **b) UC02: Publicar conjunto de dados**

**Autor principal:** Analista de BI, ou responsável designado para essa função.

**Pré-condições:** a) o ator responsável pela publicação deve estar autenticado no sistema; b) o *dataset* deve existir, com metadados obrigatórios preenchidos e pelo menos uma versão de arquivo validada.

**Referências cruzadas:** RF07, RF08, RF13, RNF01, RNF02, RNF07, RN02, RN03, RN04, RN05, RN07, RN08, RN09.

#### **Fluxo principal:**

1. O ator responsável pela publicação acessa a lista de conjuntos de dados na área administrativa.
2. Esse ator seleciona o *dataset* que deseja gerenciar.
3. O sistema apresenta o resumo do *dataset*, incluindo metadados, versões disponíveis, periodicidade de atualização e status atual.
4. O ator seleciona a opção de publicar o conjunto de dados.
5. O sistema verifica se as condições para publicação estão satisfeitas, conforme as regras de negócio.
6. O sistema atualiza o status do *dataset* para publicado.
7. O sistema passa a incluir o *dataset* nas listagens públicas e disponibiliza o arquivo tabular para *download*.
8. O sistema regista a operação no log de auditoria.
9. O caso de uso é encerrado.

#### **Fluxo alternativo: despularcar ou arquivar conjunto de dados**

1. Em vez de publicar, o ator responsável pode escolher despularcar ou arquivar um *dataset* já publicado.
2. O sistema solicita confirmação da operação.
3. Após a confirmação, o sistema atualiza o status para despularcido ou arquivado, retira o *dataset* das listagens públicas, desabilita seus painéis e regista a operação em

auditoria. 4. O fluxo continua no passo 9 do fluxo principal.

**Fluxo de exceção: condições para publicação não atendidas**

1. Caso algum requisito obrigatório, como metadados ou arquivo válido, não esteja atendido, o sistema informa que o *dataset* não pode ser publicado, indica os itens pendentes e registra a tentativa de publicação no log de auditoria. 2. O sistema mantém o status anterior do *dataset*. 3. O fluxo retorna ao passo 3 do fluxo principal.

**Pós-condição:** o *dataset* tem o status atualizado para publicado, despublicado ou arquivado, e a alteração de visibilidade está registrada em auditoria.

**c) UC03: Gerenciar visualizações analíticas**

**Autor principal:** Analista de BI.

**Atores secundários:** Plataforma de BI.

**Pré-condições:** a) o analista de BI deve estar autenticado no sistema; b) o *dataset* correspondente deve estar publicado; c) devem existir visualizações criadas na plataforma de BI a partir dos dados do *dataset*.

**Referências cruzadas:** RF09, RF10, RF11, RF13, RNF01, RNF02, RNF05, RN09.

**Fluxo principal:**

1. O analista de BI acessa a área de gestão de visualizações analíticas no portal.
2. O analista de BI seleciona o *dataset* para o qual deseja associar visualizações. 3. O sistema apresenta a lista de visualizações já registradas para o *dataset*, se houver.
4. O analista de BI seleciona a opção de cadastrar uma nova visualização. 5. O sistema solicita informações como título da visualização, tipo de painel e identificador fornecido pela plataforma de BI.
6. O analista de BI informa os dados solicitados e confirma o cadastro.
7. O sistema registra a visualização associada ao *dataset* e grava a operação em auditoria.
8. O sistema atualiza a página pública do *dataset* para incluir a nova visualização.
9. O caso de uso é encerrado.

**Fluxo de exceção 1: visualização duplicada**

1. Se o identificador informado já estiver registrado para o mesmo *dataset*, o sistema informa que a visualização já se encontra cadastrada. 2. O sistema não cria novo registro. 3. O fluxo retorna ao passo 5 do fluxo principal.

#### **Fluxo de exceção 2: erro ao validar identificador da plataforma de BI**

1. Se o sistema não conseguir validar o identificador informado, apresenta mensagem de erro ao analista de BI. 2. O sistema não registra a visualização. 3. O fluxo retorna ao passo 5 do fluxo principal.

**Pós-condição:** o *dataset* permanece publicado e passa a ter uma ou mais visualizações analíticas associadas, passíveis de exibição na interface pública.

#### **d) UC04: Consultar dados e painéis**

**Autor principal:** Usuário público.

**Atores secundários:** Plataforma de BI.

**Pré-condição:** deve haver pelo menos um *dataset* publicado no portal.

**Referências cruzadas:** RF08, RF11, RF12, RF13, RNF03, RNF04, RNF05, RN09.

#### **Fluxo principal:**

1. O usuário público acessa o portal de transparência por meio do navegador. 2. O usuário navega pelas categorias de dados ou utiliza o mecanismo de busca para localizar um *dataset* de interesse. 3. O sistema apresenta a lista de conjuntos de dados que atendem aos critérios de busca ou à categoria selecionada. 4. O usuário seleciona um *dataset* na lista. 5. O sistema exibe a página do *dataset*, incluindo metadados, opções de *download* e painéis interativos associados. 6. O usuário interage com os painéis para explorar as informações disponíveis. 7. Se desejar, o usuário solicita o *download* do arquivo tabular correspondente. 8. O sistema disponibiliza o arquivo para *download* em formato aberto. 9. O caso de uso é encerrado.

#### **Fluxo de exceção 1: nenhum *dataset* encontrado**

1. Se não houver conjuntos de dados que atendam aos critérios de busca, o

sistema informa que não foram encontrados resultados para a pesquisa realizada. 2. O sistema sugere que o usuário revise os filtros ou utilize outros termos de busca. 3. O fluxo retorna ao passo 2 do fluxo principal.

#### **Fluxo de exceção 2: falha na visualização ou no *download***

1. Se ocorrer erro na exibição dos painéis ou na disponibilização do arquivo para *download*, o sistema apresenta mensagem de indisponibilidade temporária. 2. O sistema registra a falha para posterior análise técnica. 3. O caso de uso é encerrado.

**Pós-condição:** o usuário obtém acesso às informações desejadas por meio de visualizações interativas e/ou arquivos em formato aberto, ou é informado sobre a indisponibilidade temporária dos recursos.

## **3.4 Banco de dados**

A camada de dados do portal de transparência proposto foi modelada com base no paradigma relacional, de forma a garantir organização, integridade e facilidade de consulta sobre os conjuntos de dados institucionais. Nesta seção é apresentado o modelo entidade–relacionamento adotado e são descritas as principais tabelas do banco de dados, destacando suas responsabilidades no suporte aos requisitos funcionais, às regras de negócio e aos casos de uso definidos anteriormente.

### **3.4.1 Diagrama entidade–relacionamento**

A Figura 3.2 apresenta o diagrama entidade–relacionamento (DER) do banco de dados do portal de dados e transparência. O modelo está centrado na entidade *Dataset*, que representa cada conjunto de dados institucional disponibilizado no sistema. A partir dela se organizam as demais entidades responsáveis por registrar unidades administrativas, categorias temáticas, usuários, versões de arquivos, visualizações analíticas e registros de auditoria.

A entidade *UnidadeResponsavel* armazena informações sobre as unidades organizacionais da instituição (como pró-reitorias, faculdades ou departamentos) e se

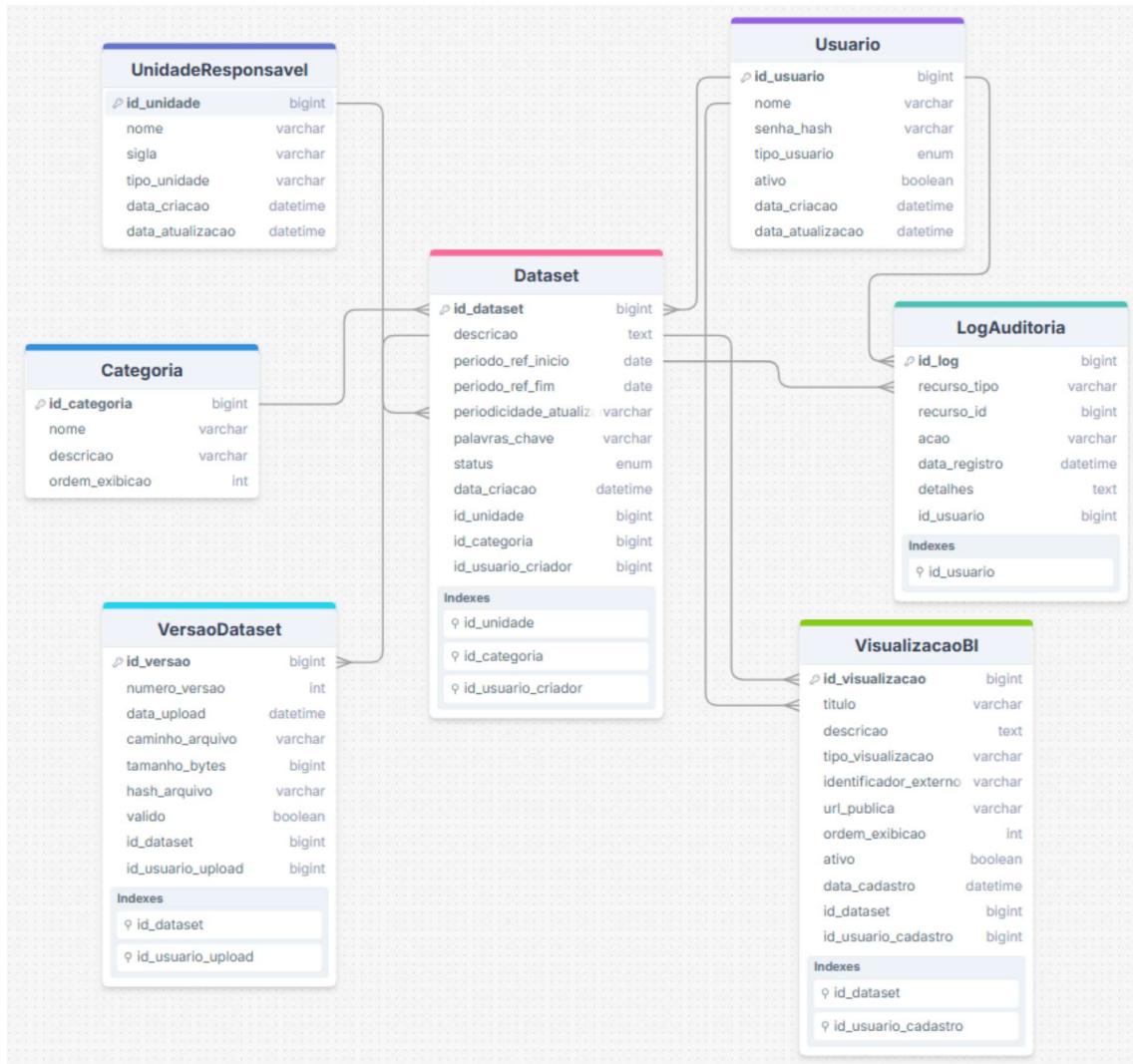


Figura 3.2: DER do banco de dados do portal de dados e transparência proposto.

relaciona com *Dataset* em uma cardinalidade de um-para-muitos, indicando que uma mesma unidade pode ser responsável por diversos conjuntos de dados. De forma análoga, a entidade *Categoria* representa agrupamentos temáticos (por exemplo, ensino, pesquisa, orçamento), também em relação um-para-muitos com *Dataset*, permitindo a classificação dos conjuntos de dados por área de interesse.

A entidade *Usuario* registra os usuários internos do sistema, responsáveis tanto pela operação administrativa do portal quanto pelo *upload* de arquivos e pelo cadastro de visualizações analíticas. Cada *Dataset* possui a referência ao usuário que realizou seu cadastro inicial, enquanto a entidade *VersaoDataset* guarda o usuário responsável pelo envio de cada versão de arquivo. A entidade *LogAuditoria*, por sua vez, mantém

o histórico das operações administrativas relevantes (como criação, publicação e arquivamento), associando cada registro a um usuário e ao recurso afetado.

O controle de versões é realizado pela entidade *VersaoDataset*, em relacionamento um-para-muitos com *Dataset*. Cada conjunto de dados pode possuir diversas versões de arquivo ao longo do tempo, preservando o histórico de atualizações e permitindo rastrear a evolução do conteúdo publicado. Já a entidade *VisualizacaoBI* representa os painéis e gráficos provenientes da plataforma de BI, associados a um determinado *Dataset* e cadastrados por um usuário com perfil adequado. Esse vínculo torna possível apresentar, em uma mesma página do portal, metadados, arquivos tabulares e visualizações interativas referentes ao mesmo conjunto de dados.

Por fim, a entidade *LogAuditoria* se relaciona com *Usuario* em uma cardinalidade um-para-muitos, registrando as ações executadas por cada usuário ao longo do tempo. Esse componente materializa os requisitos de rastreabilidade e controle previstos nas regras de negócio do portal, apoiando a governança da informação e a transparência sobre as operações realizadas no sistema.

Em conjunto, os elementos descritos neste capítulo caracterizam a proposta do sistema em nível conceitual, desde o contexto e os trabalhos relacionados até o comportamento esperado e a organização dos dados institucionais. No capítulo seguinte, são apresentadas as escolhas tecnológicas concretas que materializam essa proposta e a forma como a arquitetura implementada dá suporte às funcionalidades aqui definidas.

# Capítulo 4

## Rural.Data: Portal de Dados e Transparência da UFRRJ

Neste capítulo, descreve-se o processo de desenvolvimento da aplicação proposta, abordando as tecnologias e *frameworks* adotados para a construção da plataforma. Em seguida, discute-se como as principais funcionalidades foram implementadas, detalhando a integração entre o módulo administrativo de gestão de dados e o portal público de transparência, bem como a conexão com a ferramenta de BI.

O objetivo deste capítulo é relatar, de forma sistemática, como a solução Rural.Data foi construída, quais tecnologias e decisões de projeto foram adotadas e em que medida a implementação obtida oferece evidências de que o sistema é capaz de atender aos requisitos e objetivos definidos no Capítulo 3.

### 4.1 Tecnologias Utilizadas

A arquitetura do Rural.Data foi concebida para unir a robustez necessária no tratamento de dados com a fluidez de uma interface moderna. Para isso, selecionou-se um conjunto de ferramentas que permitisse o desenvolvimento ágil, centralizando a lógica de negócios, o que simplifica a manutenção e o *deploy* quando comparado a arquiteturas distribuídas complexas.

### 4.1.1 AdonisJS

O AdonisJS<sup>1</sup> é um *framework* Node.js robusto, construído para o desenvolvimento de aplicações web escaláveis. O AdonisJS adota uma filosofia “baterias inclusas” (*batteries-included*), o que significa que ele fornece nativamente, a partir de seu projeto-base, recursos essenciais como roteamento, sistema de autenticação, validação de dados e segurança contra ataques web comuns.

O *framework* estrutura a aplicação seguindo o padrão arquitetural *Model View Controller* (MVC), o que facilitou a organização do código do Rural.Data. O projeto-base é gerado com uma estrutura de diretórios padronizada, na qual o *framework* oferece comandos *Command Line Interface* (CLI) próprios, como o `node ace`, para a criação automatizada de *controllers*, *models* e *migrations*, garantindo que cada componente da lógica de negócios esteja em seu respectivo local.

Para o contexto deste trabalho, a escolha do AdonisJS foi determinante devido ao seu sistema eficiente de manipulação de arquivos (*BodyParser*). Como o portal exige o *upload* de grandes volumes de dados via arquivos *Comma Separated Values* (CSV), a capacidade do *framework* de processar *streams* de dados de forma nativa e segura foi um fator decisivo para sua adoção.

### 4.1.2 React

O React<sup>2</sup> é uma biblioteca JavaScript de código aberto, utilizada para construir interfaces de usuário baseadas em componentes. Sua principal característica é a abordagem declarativa, que permite aos desenvolvedores descrever como a interface deve se comportar em diferentes estados, enquanto a biblioteca se encarrega de atualizar e renderizar os componentes de forma eficiente quando os dados mudam.

No projeto Rural.Data, o React foi utilizado como a base para todo o desenvolvimento do *frontend*. Sua arquitetura baseada em componentes permitiu a modularização da interface, facilitando a manutenção e a escalabilidade do código

---

<sup>1</sup><<https://adonisjs.com>>

<sup>2</sup><<https://react.dev>>

tanto no painel administrativo quanto no portal público.

#### 4.1.3 Inertia.js

O Inertia.js<sup>3</sup> é uma biblioteca que atua como um adaptador entre *frameworks* de servidor e interfaces de cliente. Ele permite construir *Single Page Application* (SPA)s utilizando o roteamento e os *controllers* clássicos do *backend*, eliminando a complexidade de desenvolver e gerenciar uma *Application Programming Interface* (API) REST completa para a comunicação interna.

Sua utilização no sistema simplificou significativamente a arquitetura da aplicação. Ao integrar o AdonisJS com o React via Inertia, foi possível manter a lógica de navegação centralizada no servidor, enquanto o navegador renderiza os componentes dinamicamente, unindo a agilidade de desenvolvimento de um monólito à experiência de usuário de uma SPA.

#### 4.1.4 Tailwind CSS

O Tailwind CSS<sup>4</sup> é um *framework* de estilização que disponibiliza classes utilitárias de baixo nível para a construção de layouts. Ele permite aplicar estilos diretamente na marcação *HyperText Markup Language* (HTML), agilizando o processo de design e reduzindo o tamanho final dos arquivos de estilo por meio da remoção de códigos não utilizados.

No contexto do trabalho, o Tailwind CSS foi fundamental para implementar a responsividade do portal. A aplicação de suas classes utilitárias garantiu que as interfaces de visualização de dados se adaptassem corretamente a diferentes tamanhos de tela, assegurando a acessibilidade do conteúdo tanto em dispositivos móveis quanto em computadores.

---

<sup>3</sup><<https://inertiajs.com>>

<sup>4</sup><<https://tailwindcss.com>>

#### 4.1.5 ShadCN/UI

O ShadCN/UI<sup>5</sup> consiste em uma coleção de componentes de interface reutilizáveis construída sobre as primitivas do Radix UI. Diferentemente de bibliotecas convencionais instaladas via gerenciador de pacotes, ela permite que o código-fonte dos componentes seja incorporado diretamente ao projeto.

A escolha dessa ferramenta visou padronizar a identidade visual e garantir a acessibilidade do sistema. Componentes como modais, menus e formulários foram implementados seguindo padrões de acessibilidade, permitindo que a equipe focasse na implementação das funcionalidades de transparência sem a necessidade de recriar elementos básicos de interface.

#### 4.1.6 PostgreSQL

O PostgreSQL<sup>6</sup> é um Sistema Gerenciador de Banco de Dados (SGBD) de código aberto, amplamente reconhecido por sua confiabilidade e conformidade com os padrões *Structured Query Language* (SQL). Ele oferece suporte a transações Atomicidade, Consistência, Isolamento e Durabilidade (ACID) e possui recursos avançados para o armazenamento e consulta de grandes volumes de dados.

O SGBD foi adotado para atuar como o repositório central de informações do Rural.Data. A interação com o banco foi feita por meio de um *Object-Relational Mapping* (ORM), garantindo segurança nas *queries*. Ele armazena de forma estruturada tanto os dados administrativos do sistema quanto os conjuntos de dados importados via arquivos CSV, servindo também como a base de dados confiável para a conexão e geração de relatórios na ferramenta de análise.

#### 4.1.7 Docker

O Docker<sup>7</sup> é uma plataforma de software que permite a criação, teste e implantação de aplicações rapidamente por meio do uso de contêineres. Esses contêineres

---

<sup>5</sup><<https://ui.shadcn.com>>

<sup>6</sup><<https://www.postgresql.org>>

<sup>7</sup><<https://www.docker.com>>

padronizam o ambiente de execução, garantindo que o sistema funcione da mesma maneira independentemente da máquina onde está instalado.

No contexto do Rural.Data, a utilização do Docker foi determinante para a viabilidade da integração entre os componentes do sistema. A ferramenta foi a solução adotada para orquestrar a comunicação interna necessária para que a API do Metabase pudesse referenciar e ingerir corretamente as estruturas de dados (tabelas e objetos JavaScript Object Notation (*JavaScript Object Notation (JSON)*)) armazenadas no PostgreSQL. Por meio da rede interna criada pelo Docker Compose, garantiu-se que o Metabase acessasse a base de dados de forma consistente, superando limitações de conectividade e resolução de *hosts* comuns em ambientes de desenvolvimento tradicionais.

#### 4.1.8 Metabase

O Metabase<sup>8</sup> é uma plataforma de BI de código aberto que democratiza a análise de dados. A ferramenta conecta-se diretamente a bancos de dados e oferece uma interface gráfica intuitiva para a criação de consultas e *dashboards*, permitindo a visualização de métricas sem a exigência de conhecimentos avançados em SQL.

No projeto, o Metabase complementa a arquitetura ao prover a camada de visualização analítica. Ele consome os dados já persistidos no PostgreSQL por meio de uma conexão direta ao banco, e não a partir dos arquivos CSV originais. Após o processo de importação realizado pelo Rural.Data, cada arquivo é convertido em tabelas específicas no SGBD; o Metabase enxerga essas tabelas como qualquer outra fonte relacional, gerando consultas SQL automaticamente para compor gráficos dinâmicos que são incorporados ao portal público. Dessa forma, a ferramenta automatiza a transformação de dados brutos em informações visuais comprehensíveis para a comunidade acadêmica.

---

<sup>8</sup><<https://www.metabase.com>>

## 4.2 Implementação do Sistema

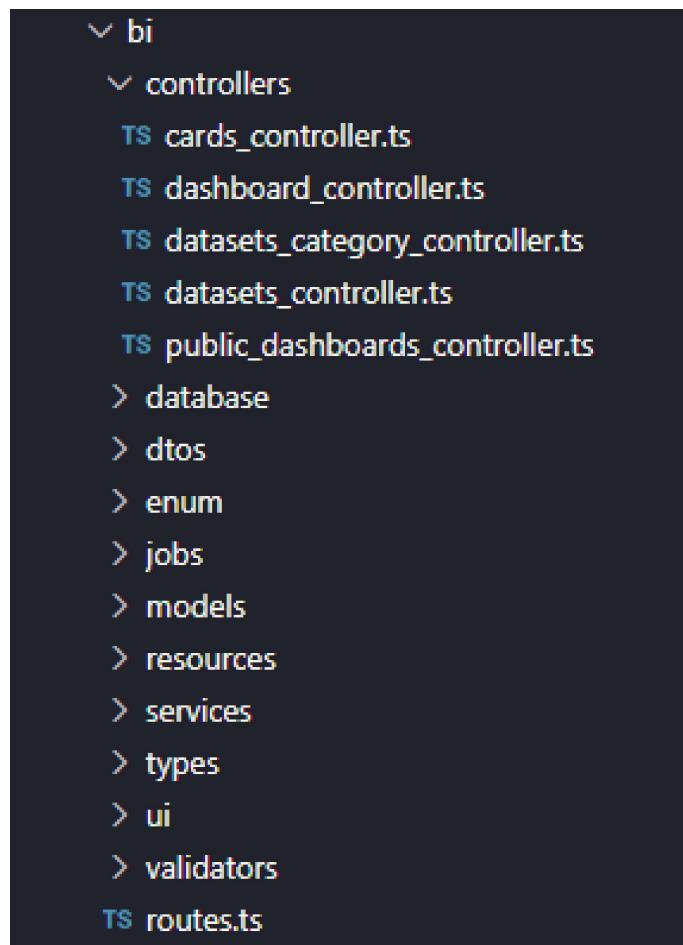
Nesta seção, detalha-se o processo de construção do Rural.Data, correlacionando as decisões arquiteturais descritas anteriormente com a implementação prática das funcionalidades. O desenvolvimento foi norteado pela necessidade de garantir a integridade dos dados institucionais durante a importação e a automação na geração de visualizações analíticas.

### 4.2.1 Estrutura de Arquivos e Organização

O desenvolvimento do sistema seguiu a estrutura modular do *framework* AdonisJS, organizando o código em diretórios semânticos para favorecer a escalabilidade e a manutenção. Para isolar as regras de negócio específicas da inteligência de dados, criou-se um módulo dedicado denominado **bi** (*Business Intelligence*), que centraliza todas as operações de manipulação de *datasets* e integração com ferramentas externas.

A Figura 4.1 ilustra a organização dos arquivos dentro deste módulo. Observa-se a separação clara das responsabilidades por meio dos diretórios: **controllers**, onde residem os arquivos responsáveis por gerenciar as requisições HTTP e a comunicação com a interface; **services**, que encapsulam a lógica de negócio complexa, como a manipulação de arquivos CSV e a comunicação com a API do Metabase; e **jobs**, destinados ao processamento assíncrono de tarefas pesadas. Além disso, o arquivo **routes.ts** define as rotas específicas deste módulo, mantendo o roteamento organizado e segregado do restante da aplicação.

Essa organização permitiu que funcionalidades críticas, como a importação de arquivos e a sincronização de painéis, fossem desacopladas da camada de apresentação, facilitando a realização de testes e a reutilização de código em diferentes partes do sistema.



```
bi
  controllers
    cards_controller.ts
    dashboard_controller.ts
    datasets_category_controller.ts
    datasets_controller.ts
    public_dashboards_controller.ts
  database
  dtos
  enum
  jobs
  models
  resources
  services
  types
  ui
  validators
  routes.ts
```

Figura 4.1: Estrutura de diretórios do módulo de *Business Intelligence*.

#### 4.2.2 Arquitetura em Camadas

Para organizar o desenvolvimento, adotou-se uma arquitetura em camadas. Esse modelo divide o sistema em partes com responsabilidades específicas, o que facilita a manutenção e a evolução do código. No Rural.Data, essa divisão permite que a lógica complexa de processamento de dados não interfira na interface do usuário.

A Figura 4.2 ilustra como essas camadas se comunicam. O fluxo é unidirecional: a interface solicita uma ação, a lógica processa e o banco de dados armazena ou recupera a informação.

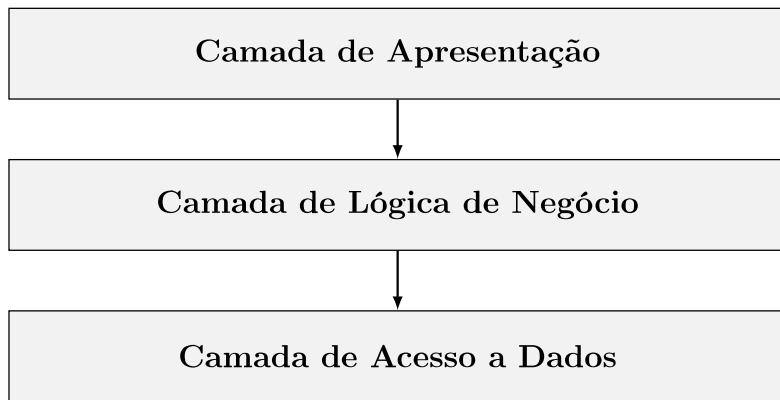


Figura 4.2: Arquitetura em camadas adotada no sistema.

#### 4.2.2.1 Camada de Apresentação

Esta camada é responsável por tudo que o usuário visualiza e com o que interage. No projeto, ela foi construída utilizando a biblioteca React dentro do ecossistema do AdonisJS, por meio do adaptador Inertia.js. Diferentemente de sistemas tradicionais que recarregam a página inteira a cada clique, o uso do Inertia permite que a navegação seja fluida (SPA), enquanto os dados são injetados diretamente nos componentes visuais pelos *controllers*.

Para agilizar o desenvolvimento e garantir a consistência visual, utilizou-se a biblioteca de componentes ShadCN/UI. Ela forneceu elementos de interface prontos e acessíveis, permitindo que o foco do desenvolvimento se mantivesse na usabilidade e na integração com os dados.

#### 4.2.2.2 Camada de Lógica de Negócio

A camada de lógica constitui o núcleo funcional da aplicação, sendo responsável pela aplicação das regras de domínio e validação de processos. Embora esta camada permeie todo o sistema, a complexidade técnica deste trabalho concentra-se no módulo de Inteligência de Negócios. Nesta arquitetura, a lógica é materializada principalmente por meio de serviços (*services*) e filas de processamento (*jobs*), que isolam o tratamento de dados das interfaces de controle.

O destaque desta implementação reside no serviço de manipulação de dados, que

executa um fluxo de processamento do tipo ETL para assegurar a integridade das informações institucionais. O processo inicia-se com as etapas de extração e transformação, em que o serviço normaliza o arquivo binário recebido. Simultaneamente, o conteúdo textual é convertido em estruturas de dados manipuláveis, permitindo que o sistema infira dinamicamente os cabeçalhos e tipos de dados.

Para reduzir o risco de interpretações incorretas, essa inferência é acompanhada de regras adicionais de validação: o serviço exige a presença de uma linha de cabeçalho, verifica se os nomes das colunas são únicos e compatíveis com o padrão de nomenclatura do banco e cruza os tipos inferidos com uma amostra mais ampla do arquivo. Caso sejam detectadas inconsistências — por exemplo, uma coluna inferida como numérica contendo valores textuais — o processamento é interrompido, o *job* é marcado como falho e o usuário responsável é notificado para revisar o arquivo e reenviar os dados.

Na etapa de carga, a lógica de negócio encarrega-se da persistência estruturada. O sistema constrói instruções SQL em tempo de execução para criar tabelas personalizadas que se adaptam exatamente à estrutura do arquivo importado. Adicionalmente, esta camada orquestra a integração com a ferramenta de BI, encapsulando a comunicação com a API externa para automatizar a criação de painéis e a publicação dos gráficos resultantes.

#### 4.2.2.3 Camada de Acesso a Dados

A camada de acesso a dados é responsável por isolar a lógica de persistência e abstrair os mecanismos de comunicação com o banco de dados, preservando a integridade das demais camadas da aplicação. Neste projeto, essa função é desempenhada pelo Lucid ORM, nativo do AdonisJS, em conjunto com o sistema gerenciador de banco de dados PostgreSQL.

O Lucid atua como uma ponte entre a aplicação e o banco, oferecendo uma interface tipada e segura para a execução de operações de leitura e escrita. A camada de dados foi projetada para criar novas tabelas em tempo de execução, modeladas com base no cabeçalho dos arquivos CSV processados pela camada lógica. Por esse motivo,

o sistema impõe requisitos mínimos de formato aos arquivos importados: é obrigatório que o CSV contenha uma primeira linha de cabeçalho, esteja codificado em UTF-8 e utilize delimitadores padronizados (vírgula ou ponto e vírgula). Arquivos que não atendem a essas condições são rejeitados ainda na etapa de validação estrutural, evitando a criação de tabelas inconsistentes no banco.

Essa abordagem permite que o sistema armazene diferentes tipos de relatórios institucionais sem a necessidade de alterações manuais no código-fonte, ao mesmo tempo em que preserva um nível mínimo de padronização sobre a estrutura dos dados recebidos.

#### 4.2.3 Portal Público e Visualização de Dados

Antes de detalhar os mecanismos restritos de gestão, é fundamental apresentar a interface de acesso público, que cumpre o objetivo primário de transparência ativa do Rural.Data. A arquitetura de informação do portal foi projetada para permitir que qualquer cidadão localize e visualize os indicadores de forma intuitiva, sem a necessidade de cadastro ou autenticação prévia.

A experiência de navegação inicia-se na página de listagem, em que o sistema apresenta um catálogo dos conjuntos de dados disponíveis. O *controller* responsável por esta rota consulta o banco de dados filtrando apenas os painéis marcados com o status “Publicado”, garantindo que rascunhos ou dados em validação não sejam expostos indevidamente. Essa tela atua como um índice organizado por categorias, permitindo ao usuário filtrar os temas de seu interesse, como “Acadêmico” ou “Orçamentário”. A Figura 4.3 ilustra essa interface de visualização.

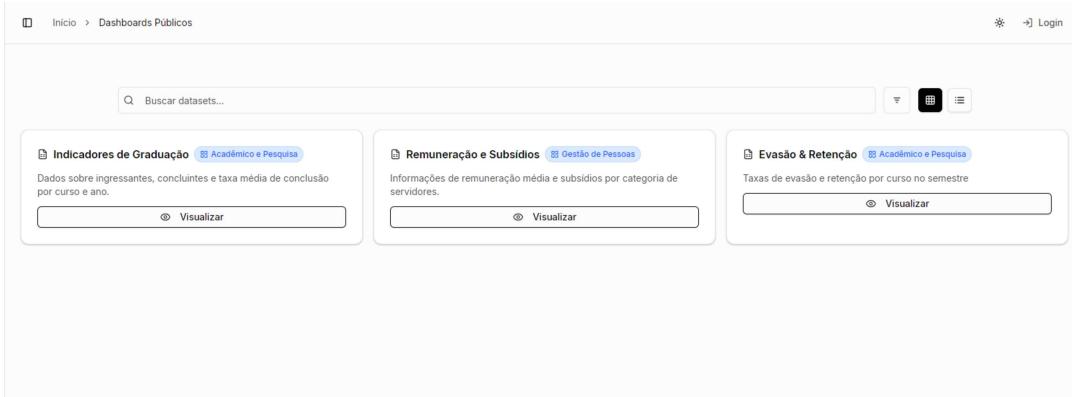


Figura 4.3: Interface pública de visualização de painel de indicadores.

Ao selecionar um dos indicadores disponíveis na listagem, o usuário é direcionado para a interface de visualização detalhada. A Figura 4.4 apresenta essa tela, em que os dados brutos são transformados em inteligência visual.

Nessa interface, a renderização dos gráficos é realizada por meio de um componente seguro que isola o contexto da aplicação do motor de processamento analítico. O sistema utiliza a técnica de *embedding* para carregar o painel correspondente ao identificador (*Universally Unique Identifier* (UUID)) selecionado.

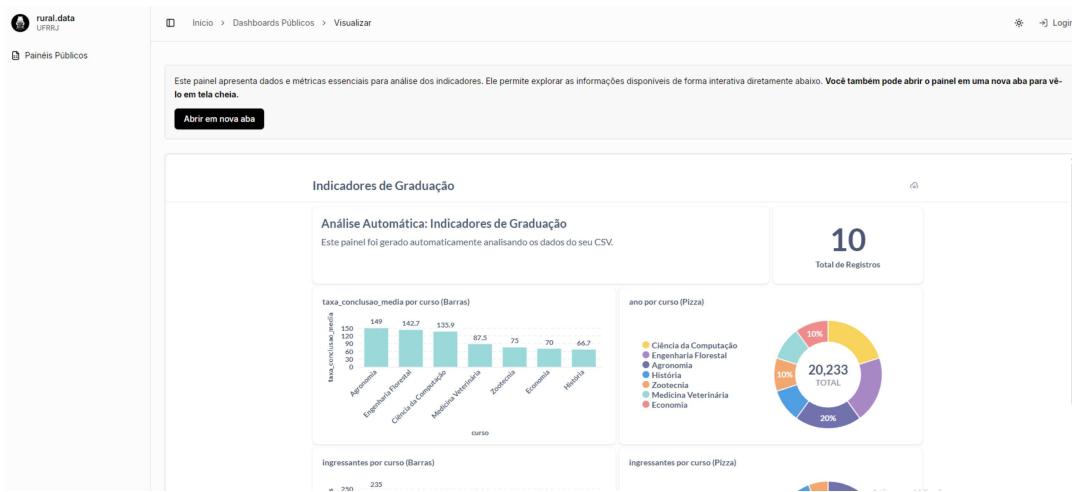


Figura 4.4: Visualização detalhada de um painel de indicadores.

Essa abordagem de segregar a listagem (índice) da visualização (detalhe) otimiza o carregamento da aplicação, pois os recursos pesados de renderização gráfica só são solicitados sob demanda, garantindo uma experiência de navegação ágil mesmo em

dispositivos móveis.

#### 4.2.4 Sistema de Autenticação e Controle de Acesso

Enquanto a área pública foca na livre distribuição de informações, a integridade dos dados depende de um controle rigoroso sobre quem pode modificá-los. Para isso, o sistema implementa um módulo de autenticação que restringe as operações de escrita exclusivamente aos gestores autorizados.

O processo de acesso administrativo inicia-se com o cadastro na plataforma, acessível por meio da rota `/signup`. A interface apresenta um formulário desenvolvido com componentes da biblioteca ShadCN/UI, projetado para garantir a usabilidade e a validação imediata dos dados inseridos. A Figura 4.5 ilustra essa tela de registro. Ao submeter as informações, o *controller* responsável aciona os serviços de validação, que verificam a unicidade do e-mail e a complexidade da senha. Para assegurar a proteção dos dados sensíveis, a senha é armazenada no banco de dados apenas após ser criptografada utilizando algoritmos de *hash* seguros.

The image shows a user registration form titled "Crie sua conta" (Create your account). The form instructions say "Preencha os campos abaixo para criar sua conta" (Fill in the fields below to create your account). It includes fields for "Nome completo" (Full name) with placeholder "Seu nome completo", "Email" with placeholder "m@ufrj.br", "Senha" (Password) with placeholder "Digite uma senha forte", "Confirme sua senha" (Confirm password) with placeholder "Repita sua senha", and a "Criar conta" (Create account) button. Below the button is a link "Já tem uma conta? Entrar" (Already have an account? Enter).

Figura 4.5: Tela de cadastro de usuário.

Após o registro, o usuário é redirecionado para a etapa de autenticação. A tela de *login* solicita as credenciais e encaminha a requisição para o serviço de autenticação. Uma vez confirmada a identidade, o sistema consulta as permissões associadas à conta. É nessa etapa que ocorre a diferenciação de perfis: caso o sistema identifique que o usuário possui privilégios administrativos atribuídos previamente por meio da infraestrutura de *backend*, a interface libera o acesso aos módulos de gestão interna.

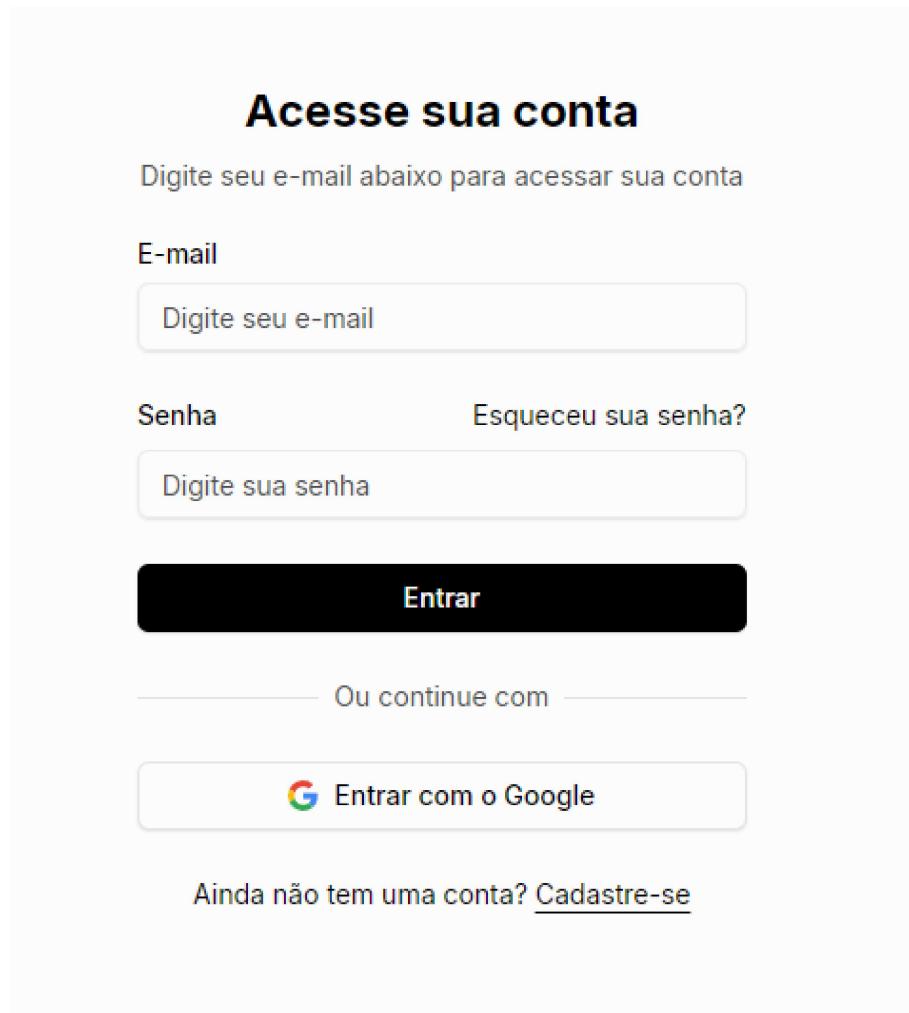


Figura 4.6: Tela de login do sistema.

Essa validação atua como um mecanismo de controle de acesso: o reconhecimento das credenciais administrativas habilita os menus de configuração e manutenção que permanecem ocultos para usuários comuns. Dessa forma, o sistema garante que as ferramentas de manipulação de *datasets*, que serão detalhadas na seção seguinte, estejam disponíveis exclusivamente para os gestores autorizados.

#### 4.2.5 Gerenciamento de Conjuntos de Dados e Painéis

O módulo de gestão de dados constitui o núcleo operacional da área restrita do sistema, sendo projetado para oferecer controle total sobre o ciclo de vida da informação publicada. Compete ao perfil de administrador a responsabilidade

de gerenciar a ingestão de arquivos brutos e a sua subsequente transformação em visualizações públicas, garantindo a curadoria dos dados antes de sua disponibilização à sociedade.

A organização interna desse módulo segue uma estrutura hierárquica em que a entidade principal é o conjunto de dados (*dataset*). A Figura 4.7 apresenta a interface de listagem e controle desses conjuntos. Nessa tela, o sistema exibe os arquivos já processados, classificados por categorias temáticas como “Acadêmico e Pesquisa” ou “Gestão de Pessoas”.

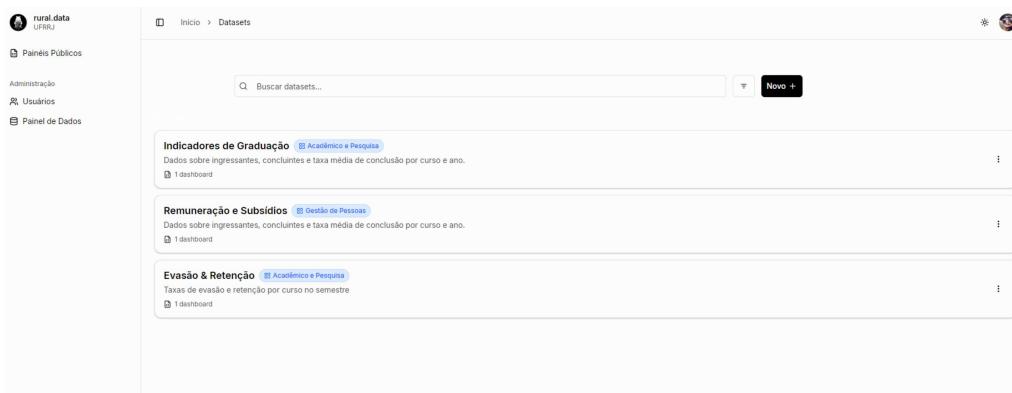
A screenshot of a web-based dataset management interface. The top navigation bar includes links for 'Início', 'Datasets', 'Novo +', and a user profile icon. On the left, a sidebar menu lists 'Painéis Públicos', 'Administração', 'Usuários', and 'Painel de Dados'. The main content area displays three dataset cards: 'Indicadores de Graduação' (Acadêmico e Pesquisa), 'Remuneração e Subsídios' (Gestão de Pessoas), and 'Evasão & Retenção' (Acadêmico e Pesquisa). Each card includes a brief description and a '1 dashboard' link.

Figura 4.7: Painel de gestão de conjuntos de dados (*datasets*).

Ao acionar a criação de um novo registro, o sistema apresenta o formulário de importação, ilustrado na Figura 4.8. Esta interface foi desenhada para simplificar o processo de carga: o administrador define o nome, a categoria temática e seleciona o arquivo CSV local. O envio deste formulário atua como o gatilho para o *pipeline* de ETL descrito na camada de lógica, iniciando o processamento assíncrono que converte o arquivo estático em tabelas consultáveis no banco de dados.

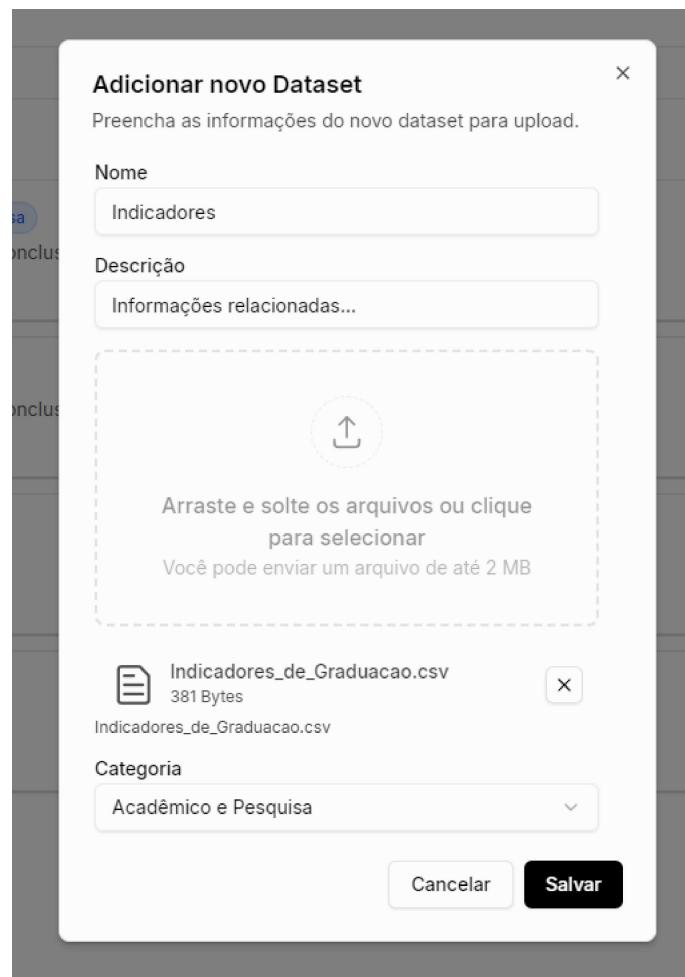


Figura 4.8: Formulário de importação de novo conjunto de dados.

Uma vez processado o conjunto de dados, o sistema habilita o nível seguinte da hierarquia: os painéis de visualização (*dashboards*). A arquitetura do sistema estabelece um relacionamento de um-para-muitos, permitindo que um único arquivo de dados alimente múltiplas visualizações distintas. A Figura 4.9 ilustra a interface de gerenciamento desses painéis, em que o administrador dispõe de ferramentas para criar novas análises gráficas, editar configurações existentes e monitorar o status de visibilidade de cada item.

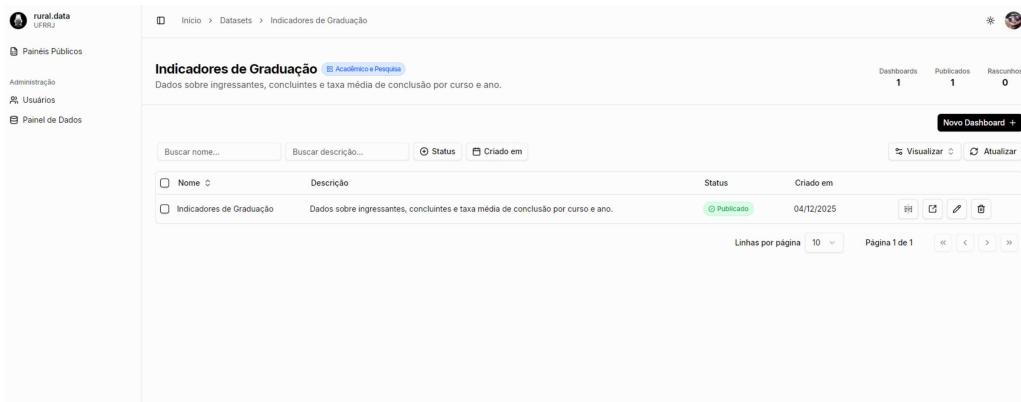


Figura 4.9: Interface de controle de painéis vinculados a um *dataset*.

Um aspecto crítico implementado nessa interface é o controle de publicação. O sistema adota um modelo de estados para os painéis, distinguindo entre “Rascunho” e “Publicado”. Essa distinção assegura que o processo de criação e validação dos gráficos ocorra em ambiente isolado. Apenas mediante uma ação explícita de publicação por parte do administrador é que o sistema gera os identificadores seguros e libera o acesso ao conteúdo na interface pública.

A implementação detalhada ao longo deste capítulo demonstra que a arquitetura proposta para o Rural.Data é capaz de solucionar o problema da dispersão e dificuldade de acesso aos dados institucionais. Ao integrar um fluxo robusto de ingestão dinâmica (ETL) com uma camada de visualização automatizada (BI), o sistema atende integralmente aos requisitos funcionais e não funcionais estabelecidos, garantindo a integridade das informações durante o processamento e a alta disponibilidade na consulta pública. Dessa forma, a solução materializa-se não apenas como um repositório de arquivos, mas como uma ferramenta efetiva de transparência ativa, validando a viabilidade técnica de modernizar a divulgação de dados públicos por meio de uma infraestrutura moderna e escalável.

# Capítulo 5

## Considerações Finais

Este capítulo apresenta as considerações finais deste trabalho, retomando o problema investigado, sintetizando o percurso desenvolvido ao longo dos capítulos anteriores, destacando as contribuições obtidas e apontando limitações e possibilidades de evolução da solução proposta.

### 5.1 Síntese do trabalho desenvolvido

O ponto de partida deste estudo, apresentado no Capítulo 1, foi a constatação de que, embora a agenda de dados abertos e transparência tenha avançado no âmbito da administração pública e das instituições federais de ensino superior, ainda persistem desafios significativos relacionados à organização, à publicação e ao uso estratégico dos dados institucionais. Em particular, observou-se um descompasso entre, de um lado, portais de dados abertos que se concentram na oferta de arquivos para *download* e, de outro, soluções de BI frequentemente restritas a ambientes internos, com pouco diálogo direto com os mecanismos de transparência ativa.

Nesse contexto, definiu-se como objetivo geral conceber e desenvolver um portal de dados e transparência para a Universidade Federal Rural do Rio de Janeiro, integrando a publicação de conjuntos de dados institucionais e visualizações analíticas baseadas em conceitos de BI. Para atingir esse objetivo, o trabalho foi estruturado em etapas,

cada uma delas tratada em um capítulo específico.

No Capítulo 2, foram discutidos os fundamentos teóricos relacionados à análise de dados e BI, incluindo conceitos, arquiteturas analíticas, processos de ETL e aplicações em contextos universitários. Essa fundamentação permitiu situar o papel de *data warehouses*, *data lakes* e plataformas de visualização no suporte à tomada de decisão, bem como reconhecer o potencial dessas tecnologias para organizar e monitorar informações acadêmicas e administrativas.

No Capítulo 3, foi apresentada a proposta conceitual do sistema. Foram descritos o contexto motivador, os trabalhos relacionados em portais de dados abertos de universidades federais, a visão geral do desenho do sistema, os atores envolvidos, os requisitos funcionais e não funcionais, as regras de negócio e os principais casos de uso. Essa especificação delineou uma solução centrada no conceito de *dataset* institucional, organizada em camadas de gestão interna, serviços de dados e publicação e análise, com controle explícito de versões e integração sistemática com uma plataforma de BI.

No Capítulo 4, a proposta foi materializada na forma da aplicação Rural.Data. Foram descritas as tecnologias empregadas (AdonisJS, React, Inertia.js, Tailwind CSS, ShadCN/UI, PostgreSQL, Docker e Metabase), as decisões arquiteturais que levaram à adoção de uma aplicação monolítica em camadas, os mecanismos de autenticação e controle de acesso e o fluxo de implementação das principais funcionalidades. Em seguida, detalharam-se a estrutura de arquivos, o módulo dedicado à inteligência de dados, o *pipeline* de ETL para ingestão de arquivos CSV, o repositório relacional de dados, a integração com o Metabase e as interfaces de portal público e de gestão interna de conjuntos de dados e painéis.

Em conjunto, esses elementos indicam que o objetivo de conceber e implementar um protótipo funcional de portal de dados e transparência, articulando catálogo de *datasets* institucionais e camada analítica de BI, foi alcançado dentro dos limites e do escopo de um trabalho de conclusão de curso.

## 5.2 Contribuições do trabalho

As contribuições deste trabalho podem ser analisadas em três dimensões principais: conceitual, técnica e institucional.

Do ponto de vista conceitual, o trabalho consolida uma visão integrada para portais de dados e transparência em instituições de ensino superior, ao articular de forma explícita três dimensões que frequentemente aparecem dissociadas: (i) a organização da informação em conjuntos de dados institucionais, com metadados, unidades responsáveis e ciclos de atualização; (ii) a disponibilização desses dados em formatos abertos, com mecanismos de busca, filtragem e *download*; e (iii) a associação sistemática com visualizações analíticas e painéis interativos gerados em uma plataforma de BI. Essa visão contribui para superar a separação tradicional entre portais de dados abertos e ambientes internos de análise.

Em termos técnicos, o Rural.Data demonstra a viabilidade de implementar essa integração utilizando tecnologias de código aberto amplamente adotadas. A arquitetura em camadas baseada em AdonisJS, React e Inertia.js, combinada com o uso de PostgreSQL como repositório central e Metabase como motor analítico, resultou em um protótipo capaz de receber arquivos CSV contendo dados institucionais, validar sua estrutura e persistir seu conteúdo em tabelas geradas dinamicamente; manter o controle de versões de cada *dataset*, preservando o histórico de atualizações; expor um catálogo público de conjuntos de dados com metadados descritivos, filtros temáticos e opções de *download* em formato aberto; associar a cada *dataset* um ou mais painéis analíticos criados no Metabase e incorporados por meio de *embedding* seguro; e diferenciar claramente as rotinas de acesso público das operações administrativas de cadastro, importação, publicação, despublicação e arquivamento de *datasets*.

Essas características respondem aos requisitos funcionais e não funcionais delineados no Capítulo 3 e evidenciam que a abordagem proposta é tecnicamente factível, mesmo em ambientes com recursos limitados, como é comum em instituições públicas de ensino superior.

Na dimensão institucional, ainda que o escopo deste trabalho tenha se limitado à

construção de um protótipo, a solução desenvolvida oferece um modelo concreto que pode subsidiar discussões internas sobre a modernização das práticas de transparência e de gestão baseada em dados. Ao alinhar-se à Política de Dados Abertos do Poder Executivo federal e ao Plano de Dados Abertos da UFRRJ, o Rural.Data pode ser entendido como um artefato de referência para futuras iniciativas que visem consolidar um portal institucional de dados conectando, de maneira mais estreita, produção de dados, governança da informação e uso analítico em apoio à gestão.

### 5.3 Limitações e trabalhos futuros

Como todo trabalho de conclusão de curso, a solução apresentada possui limitações decorrentes de decisões de escopo, de tempo disponível para desenvolvimento e de restrições técnicas.

Uma primeira limitação refere-se à abrangência dos dados utilizados. O protótipo foi exercitado com um conjunto reduzido de bases institucionais, suficientes para demonstrar o funcionamento do *pipeline* de ingestão e da camada de visualização, mas ainda distantes da diversidade de fontes, formatos e volumes de dados presentes na rotina completa da universidade. A ampliação do número de *datasets* e a inclusão de bases com maior complexidade e criticidade constituem um passo importante para a consolidação da solução em ambiente real.

Outra limitação diz respeito à integração com a infraestrutura institucional de autenticação e autorização. No estágio atual, o sistema utiliza um mecanismo próprio de cadastro e *login*, adequado para fins de prototipagem, mas que não reflete o cenário desejável em produção, no qual se esperaria integração com sistemas de identidade corporativos e com perfis de acesso já estabelecidos na universidade.

Do ponto de vista da governança de dados, embora o trabalho tenha definido regras de negócio para metadados obrigatórios, controle de versões, tratamento de dados sensíveis e periodicidade de atualização, a operacionalização plena dessas regras exige processos institucionais complementares, como a definição formal de unidades responsáveis, fluxos de validação e mecanismos de revisão periódica dos

*datasets*. Esses aspectos extrapolam o escopo técnico do protótipo, mas são essenciais para que o portal contribua, de fato, para a transparência e a confiabilidade da informação publicada.

Além disso, a solução atual concentra-se na ingestão de dados por meio de arquivos tabulares e no consumo direto pelo Metabase. Uma linha promissora de evolução consiste em incorporar mecanismos adicionais de acesso automatizado, como APIs abertas para consulta programática dos *datasets*, e em explorar a integração com outras ferramentas analíticas e de ciência de dados, ampliando o ecossistema de reutilização das informações.

Diante dessas limitações, delineiam-se alguns caminhos para trabalhos futuros. Em primeiro lugar, propõe-se ampliar o conjunto de *datasets* institucionais integrados ao portal, contemplando diferentes áreas como acadêmica, orçamentária, de pessoal e assistência estudantil e volumes maiores de dados, de modo a aproximar o protótipo de um cenário operacional mais próximo da realidade da universidade. Também se considera fundamental integrar o mecanismo de autenticação do Rural.Data a um provedor de identidade institucional, com suporte a *Single Sign-On* (SSO) e a perfis de acesso alinhados à estrutura organizacional da UFRRJ, reduzindo redundâncias e aumentando a segurança no controle de permissões.

Outro eixo de avanço envolve o amadurecimento dos mecanismos de governança de dados associados ao portal. Em parceria com as unidades administrativas, sugere-se formalizar fluxos de validação, revisão e atualização de *datasets*, estabelecendo responsabilidades claras e rotinas de acompanhamento que assegurem a atualidade e a confiabilidade das informações disponibilizadas. Em paralelo, a disponibilização de interfaces de acesso automatizado, por meio de APIs abertas aos conjuntos de dados publicados, pode favorecer o desenvolvimento de aplicações externas, bem como a realização de análises por pesquisadores e demais interessados, ampliando o potencial de reuso das informações.

Por fim, recomenda-se explorar novas formas de visualização e interação com os dados, incluindo narrativas visuais, painéis temáticos e recursos que facilitem a compreensão por públicos não especializados, e avaliar o uso do Rural.Data em

estudos de caso ampliados, com participação de usuários internos e externos. Esses estudos podem fornecer insumos valiosos sobre usabilidade, utilidade percebida e impacto na transparência da informação, orientando ciclos sucessivos de melhoria da solução.

## 5.4 Considerações finais

Ao longo deste trabalho, buscou-se responder ao desafio de aproximar dados abertos e inteligência de negócios em um contexto universitário específico, propondo e implementando um portal de dados e transparência capaz de integrar catálogo de *datasets* institucionais, controle de versões, metadados descritivos e painéis analíticos interativos.

Ainda que se trate de um protótipo, o Rural.Data demonstra que é possível superar modelos centrados apenas na disponibilização de arquivos, avançando em direção a uma transparência ativa em que dados brutos, contexto informacional e visualizações interpretáveis coexistem em uma mesma solução. Espera-se que as contribuições apresentadas possam servir de base para iniciativas futuras na UFRRJ e em outras instituições de ensino superior, colaborando para o fortalecimento de uma cultura de gestão baseada em dados e para a consolidação de práticas de abertura e reuso de informações públicas em benefício da sociedade.

# Referências

- ALBANO, C. S.; CRAVEIRO, G. d. S.; GAFFRÉE, J. R. d. L. Oferta de dados abertos em universidades federais brasileiras: um estudo dos planos de dados abertos. *Acervo*, v. 34, n. 3, p. 1–18, 2021.
- ARMBRUST, M. et al. Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. In: *Proceedings of the 11th Annual Conference on Innovative Data Systems Research (CIDR '20)*. [S.l.]: [s.n.], 2020.
- BRASIL. *Decreto nº 8.777, de 11 de maio de 2016. Institui a Política de Dados Abertos do Poder Executivo federal*. 2016. <<http://www.planalto.gov.br/>>. Diário Oficial da União, Brasília, DF. Acesso em: 05 mar. 2025.
- BRASIL. *Portal Brasileiro de Dados Abertos*. 2025. <<https://dados.gov.br/>>. Acesso em: 07 mar. 2025.
- CHAUDHURI, S.; DAYAL, U.; NARASAYYA, V. An overview of business intelligence technology. *Communications of the ACM*, v. 54, n. 8, p. 88–98, 2011.
- INMON, W. H. *Building the Data Warehouse*. 4. ed. Hoboken: John Wiley & Sons, 2005.
- JÚNIOR, O. d. G. F. et al. Uma experiência com *Business Intelligence* para apoiar a gestão acadêmica em uma universidade federal brasileira. *Revista Ibérica de Sistemas e Tecnologias de Informação*, n. 46, p. 5–20, 2022.
- KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3. ed. Indianapolis: Wiley, 2013.
- LIMA, M. S.; PIRES, M. R. G. M. Avaliação da taxa do acesso aos dados abertos das universidades federais a partir dos indicadores de fluxo do ensino superior do inep. *Avaliação: Revista da Avaliação da Educação Superior*, Campinas; Sorocaba, SP, v. 27, n. 3, p. 531–552, 2022. Acesso em: 10 abr. 2025. Disponível em: <<https://www.scielo.br/j/aval/a/hSWjCWMY5GmRgp5pZNQ5Kys/?lang=pt>>.
- LOSHIN, D. *Business Intelligence: The Savvy Manager's Guide*. Amsterdam: Elsevier, 2012.
- MOSS, L. T.; ATRE, S. *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*. [S.l.]: Addison-Wesley, 2003.

- PIRES, K. M. Dados abertos nas universidades federais: envolvimento interno e divulgação para a sociedade. *Revista Brasileira de Biblioteconomia e Documentação*, v. 15, n. 2, p. 58–80, 2019.
- SAWADOGO, P. N.; DARMONT, J. On data lake architectures and metadata management. *Journal of Intelligent Information Systems*, v. 56, n. 1, p. 97–120, 2021.
- TURBAN, E.; SHARDA, R.; DELEN, D. *Business Intelligence and Analytics: Systems for Decision Support*. 10. ed. [S.l.]: Pearson Education Limited, 2014.
- UFAC. *Plano de Dados Abertos da Universidade Federal do Acre*. 2024. <<https://www.ufac.br/>>. Acesso em: 10 abr. 2025.
- UFC. *Plano de Dados Abertos da Universidade Federal do Ceará*. 2024. <<https://www.ufc.br/>>. Acesso em: 05 abr. 2025.
- UFCA. *Plano de Dados Abertos da Universidade Federal do Cariri*. 2025. <<https://acessoainformacao.ufca.edu.br/>>. Acesso em: 25 mar. 2025.
- UFCA. *Portal de Dados Abertos da Universidade Federal do Cariri*. 2025. <<https://dados.ufca.edu.br/>>. Acesso em: 22 mar. 2025.
- UFG. *Plano de Dados Abertos da Universidade Federal de Goiás*. 2023. <<https://seti.ufg.br/>>. Acesso em: 12 mar. 2025.
- UFG. *Portal de Dados Abertos da Universidade Federal de Goiás*. 2025. <<https://dados.ufg.br/>>. Acesso em: 10 mar. 2025.
- UFRRJ. *Plano de Dados Abertos da UFRRJ 2020–2022*. 2020. <<https://portal.ufrrj.br/>>. Acesso em: 08 abr. 2025.
- UFRRJ. *Relatório de monitoramento do Plano de Dados Abertos da UFRRJ – 2023*. 2023. <<https://portal.ufrrj.br/>>. Acesso em: 08 abr. 2025.
- UFU. *Plano de Dados Abertos da Universidade Federal de Uberlândia*. 2023. <<https://www.comunica.ufu.br/>>. Acesso em: 18 mar. 2025.
- UFU. *Portal de Dados Abertos da Universidade Federal de Uberlândia*. 2025. <<https://dados.ufu.br/>>. Acesso em: 15 mar. 2025.
- VIEIRA, A. R. R.; SEABRA, R. D.; GARCÉS, L. Business intelligence for academic management systems: Results of a study in a brazilian federal university. In: *Anais do 21º Simpósio Brasileiro de Sistemas de Informação (SBSI)*. Porto Alegre: Sociedade Brasileira de Computação, 2025. p. 221–230.