



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA

LUIZA LIMA MEIRA DE MENEZES

**Implementação de Portais de Dados Abertos em Instituições Federais de Ensino Superior:** Um Estudo de Caso com a Universidade Federal Rural de Pernambuco

Recife

2023

LUIZA LIMA MEIRA DE MENEZES

**Implementação de Portais de Dados Abertos em Instituições Federais de Ensino**

**Superior:** Um Estudo de Caso com a Universidade Federal Rural de Pernambuco

Dissertação apresentada ao Programa de Pós-graduação em Informática Aplicada da Universidade Federal Rural de Pernambuco como exigência parcial à obtenção do título de Mestre em Informática Aplicada.

**Área de Concentração:** Computação Inteligente e Modelagem

**Orientador (a):** Profa. Dra. Maria da Conceição Moraes Batista

**Coorientador (a):** Profa. Dra. Roberta Macêdo Marques Gouveia

Recife

2023

Dados Internacionais de Catalogação na Publicação  
Universidade Federal Rural de Pernambuco  
Sistema Integrado de Bibliotecas  
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

---

- M543ii Menezes, Luiza Lima Meira de  
Implementação de Portais de Dados Abertos em Instituições Federais de Ensino Superior: Um Estudo de Caso com a Universidade Federal Rural de Pernambuco / Luiza Lima Meira de Menezes. - 2023.  
82 f.
- Orientadora: Maria da Conceicao Moraes Batista.  
Coorientadora: Roberta Macedo Marques Gouveia.  
Inclui referências e apêndice(s).
- Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Informática Aplicada, Recife, 2023.
1. dados abertos. 2. dados abertos governamentais. 3. lei de acesso à informação. I. Batista, Maria da Conceicao Moraes, orient. II. Gouveia, Roberta Macedo Marques, coorient. III. Título

---

CDD 004

Dissertação elaborada por Luiza Lima Meira de Menezes, intitulada Implementação de Portais de Dados Abertos em Instituições Federais de Ensino Superior: Um Estudo de Caso com a Universidade Federal Rural de Pernambuco, apresentada ao Programa de Pós-graduação em Informática Aplicada da Universidade Federal Rural de Pernambuco para obtenção do título de Mestre em Informática Aplicada, defendida e aprovada por unanimidade em 31 de março de 2023 pela comissão examinadora.

Banca Examinadora:

---

Dra. MARIA DA CONCEIÇÃO MORAES BATISTA  
Presidente (DEINFO/UFRPE)

---

Dra. ROBERTA MACÊDO MARQUES GOUVEIA  
Examinadora Externa ao Programa (DC/UFRPE)

---

Dr. GABRIEL ALVES DE ALBUQUERQUE JUNIOR  
Examinador Interno (DEINFO/UFRPE)

---

Dra. TACIANA PONTUAL DA ROCHA FALCÃO  
Examinadora Interna (DEINFO/UFRPE)

## **AGRADECIMENTOS**

Agradeço primeiramente às minhas orientadoras, Ceça Moraes e Roberta Macêdo, pela confiança, parceria e dedicação ao longo da elaboração dessa pesquisa.

Agradeço também aos professores do PPGIA, pelo conhecimento e inspiração, e aos meus colegas de turma, pelo companheirismo e apoio.

Aos membros do CTDA, que me receberam tão bem nesse projeto, e aos setores da UFRPE que contribuíram para a realização da pesquisa: PREG, PROPLAN, GCF e CATF.

Aos meus colegas da STD, por apoiar minha jornada acadêmica e pela ajuda nos aspectos de configuração de servidor e redes durante a implantação do portal de dados abertos.

## RESUMO

As iniciativas de Governo Aberto e a Lei de Acesso à Informação (LAI) deram início a um processo de mudança na administração pública, visando a transparência como forma de participação social. Uma das formas de atingir esse objetivo consiste em desenvolver e manter um portal de dados abertos, no qual serão disponibilizadas as informações de interesse da sociedade definidas pelo Plano de Dados Abertos (PDA) do órgão ou instituição. No entanto, as instituições enfrentam uma série de desafios para implementar tal portal. Entre eles, pode-se destacar a escolha da aplicação de gerenciamento de conteúdo, desconhecimento dos requisitos que tal sistema precisa possuir e falta de um guia de boas práticas para extração, limpeza e tratamento das informações que serão publicadas, para que estas possam ser consideradas abertas. Essa pesquisa tem como principal objetivo mapear as melhores tecnologias para desenvolver e gerenciar um portal de dados abertos em instituições de ensino público, assim como estabelecer um procedimento para extração, manipulação e publicação dos dados. Como prova de conceito, foi realizado um estudo de caso com as bases de Ensino de Graduação, Contratos e Orçamento da Universidade Federal Rural de Pernambuco (UFRPE), resultando na implantação do portal de dados abertos da instituição.

**Palavras-chaves:** Dados Abertos, Dados Abertos Governamentais, CKAN, Lei de Acesso à Informação.

## ABSTRACT

The Open Government Data initiative and the Access to Information Law marked the beginning of a change process in public administration, with transparency as a way of social participation. One way to achieve this goal is to develop and maintain an open data portal containing information pertinent to civil society as defined by the Open Data Plan (PDA) of the institution. However, government bodies face a number of challenges in implementing such a portal. Among them, one can highlight the choice of content management application, lack of knowledge of the requirements of such a system and the absence of a guide for extraction, cleaning and treatment of the information that will be published, so that it can be considered open data. The main goal of this research is to chart the best technologies to develop and manage an open data portal for public education institutions and establish a procedure for extracting, manipulating and publishing such data. As proof of concept, a case study was carried out with the Graduation, Contracts and Budget datasets of the Universidade Federal Rural de Pernambuco (UFRPE), resulting in the implementation of the institution's open data portal.

**Keywords:** Open Data, Open Government Data, Access to Information Law, CKAN.

## LISTA DE FIGURAS

Figura 1 – Painel de Monitoramento de Dados Abertos - Bases Previstas . . . . .	16
Figura 2 – Geração de valor a partir de dados abertos governamentais . . . . .	21
Figura 3 – Princípios da LGPD . . . . .	26
Figura 4 – Fluxo de informação . . . . .	30
Figura 5 – Adesão dos portais de transparência aos princípios de OGD . . . . .	32
Figura 6 – Qualidade dos conjuntos de dados . . . . .	34
Figura 7 – Estrutura do CKAN . . . . .	44
Figura 8 – Página Inicial do portal . . . . .	45
Figura 9 – Utilização do VLibras no CKAN . . . . .	46
Figura 10 – Página Inicial do portal após personalização . . . . .	47
Figura 11 – Plugin para visualização de dados em tabela . . . . .	48
Figura 12 – Plugin para visualização de dados como gráfico . . . . .	48
Figura 13 – Plugin para visualização de documentos PDF . . . . .	49
Figura 14 – Agrupamento dos conjuntos de dados . . . . .	50
Figura 15 – Organização de grupos no portal . . . . .	51
Figura 16 – Página da UFRPE no portal GOV.BR . . . . .	52
Figura 17 – Criação de função auxiliar para filtro de cursos ativos . . . . .	57
Figura 18 – Pseudocódigo de consulta para extração de dados de graduação . . . . .	58
Figura 19 – Conjunto de dados de Turmas de Graduação no CKAN . . . . .	58
Figura 20 – Pseudocódigo de consulta para quantitativo de alunos . . . . .	59
Figura 21 – Quantitativo de alunos de graduação no CKAN . . . . .	59
Figura 22 – Pseudocódigo de consulta para quantitativo de matrículas . . . . .	60
Figura 23 – Quantitativos de matriculados nas turmas de graduação no CKAN . . . . .	60
Figura 24 – Dados do Censo em visualização no CKAN . . . . .	62
Figura 25 – Painel de execução orçamentária da GCF . . . . .	63
Figura 26 – Grupo Orçamento e Financeiro no Portal . . . . .	64
Figura 27 – Base de Contratos com metadados no CKAN . . . . .	66
Figura 28 – Processo de publicação - Preparação . . . . .	68
Figura 29 – Aplicação da licença PDDL no CKAN . . . . .	70
Figura 30 – Justificativa para uso de dado pessoal conforme LGPD . . . . .	70

Figura 31 – Processo de publicação - Metadados . . . . .	71
Figura 32 – Processo de publicação de dados abertos . . . . .	72
Figura 33 – Papéis na publicação de dados abertos . . . . .	73
Figura 34 – Adicionar conjunto de dados no CKAN . . . . .	74
Figura 35 – Campos do conjunto de dados no CKAN . . . . .	75
Figura 36 – Gerenciar conjunto de dados no CKAN . . . . .	75

## LISTA DE QUADROS

Quadro 1 – Diferenças entre transparência ativa e passiva . . . . .	23
Quadro 2 – Comparativo de licenças abertas . . . . .	28
Quadro 3 – Matriz de priorização do PDA da UFRPE . . . . .	29
Quadro 4 – Termos do VCGE para as bases do PDA . . . . .	39
Quadro 5 – Requisitos negociais para portais de dados abertos . . . . .	39
Quadro 6 – Requisitos técnicos para portais de dados abertos . . . . .	42
Quadro 7 – Serviços que compõe o CKAN . . . . .	44
Quadro 8 – Lista de <i>plugins</i> instalados . . . . .	47

## LISTA DE ABREVIATURAS E SIGLAS

<b>API</b>	Interface de Programação de Aplicação
<b>ASF</b>	Apache Software Foundation
<b>BC</b>	Banco Central do Brasil
<b>CAPTCHA</b>	Completely Automated Public Turing test to tell Computers and Humans Apart
<b>CATF</b>	Coordenadoria de Acompanhamento Técnico Financeiro
<b>CC</b>	Creative Commons
<b>CC BY 4.0</b>	Creative Commons Atribuição 4.0
<b>CC0</b>	Creative Commons Zero
<b>CGINDA</b>	Comitê Gestor da INDA
<b>CGU</b>	Controladoria-Geral da União
<b>CKAN</b>	Comprehensive Knowledge Archive Network
<b>CONSU</b>	Conselho Universitário
<b>CPF</b>	Cadastro de Pessoa Física
<b>CSS</b>	Cascading Style Sheets
<b>CSV</b>	Comma-Separated Values
<b>CTDA</b>	Comitê de Transparência e Dados Abertos
<b>DDoS</b>	Distributed Denial of Service
<b>e-MAG</b>	Modelo de Acessibilidade em Governo Eletrônico
<b>e-PING</b>	Padrões de Interoperabilidade de Governo Eletrônico
<b>e-SIC</b>	Sistema Eletrônico do Serviço de Informações ao Cidadão
<b>ETL</b>	Extraction-Transformation-Loading
<b>Fala.BR</b>	Plataforma Integrada de Ouvidoria e Acesso à Informação
<b>GCF</b>	Gerência de Contabilidade e Finanças
<b>GDPR</b>	Regulamento Geral sobre Proteção de Dados

<b>GML</b>	Geography Markup Language
<b>GTA</b>	Guia de Transparência Ativa
<b>HTML</b>	HyperText Markup Language
<b>IFES</b>	Instituições Federais de Ensino Superior
<b>INDA</b>	Infraestrutura Nacional de Dados Abertos
<b>INDE</b>	Infraestrutura Nacional de Dados Espaciais
<b>INEP</b>	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
<b>JSON</b>	Notação de Objetos JavaScript
<b>LAI</b>	Lei de Acesso à Informação
<b>LGPD</b>	Lei Geral de Proteção de Dados Pessoais
<b>LIBRAS</b>	Língua Brasileira de Sinais
<b>MB</b>	Megabytes
<b>MEC</b>	Ministério da Educação
<b>ODbL</b>	Open Data Commons Open Database License
<b>ODP</b>	Open Data Platform
<b>ODS</b>	Open Document Spreadsheet
<b>OGD</b>	Open Government Data
<b>OKFN</b>	Open Knowledge Foundation
<b>ONG</b>	Organização Não Governamental
<b>PDA</b>	Plano de Dados Abertos
<b>PDDL</b>	Public Domain Dedication and License
<b>PDF</b>	Portable Document Format
<b>PREG</b>	Pró-Reitoria de Ensino de Graduação
<b>PROAD</b>	Pró-Reitoria de Administração
<b>PROPLAN</b>	Pró-Reitoria de Planejamento
<b>RDF</b>	Resource Description Framework
<b>RG</b>	Registro Geral

<b>SGBD</b>	Sistema de Gerenciamento de Banco de Dados
<b>SIGAA</b>	Sistema Integrado de Gestão de Atividades Acadêmicas
<b>SIGs</b>	Sistemas Institucionais Integrados de Gestão
<b>SMTP</b>	Simple Mail Transfer Protocol
<b>SQL</b>	Structured Query Language
<b>STD</b>	Secretaria de Tecnologias Digitais
<b>SVG</b>	Scalable Vector Graphics
<b>TI</b>	Tecnologia de Informação
<b>TIC</b>	Tecnologia de Informação e Comunicação
<b>UFRN</b>	Universidade Federal do Rio Grande do Norte
<b>UFRPE</b>	Universidade Federal Rural de Pernambuco
<b>URL</b>	Uniform Resource Locator
<b>VCGE</b>	Vocabulário Controlado do Governo Eletrônico
<b>VM</b>	Máquina Virtual
<b>WEB</b>	World Wide Web
<b>XML</b>	Extensible Markup Language

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
<b>1.1</b>	<b>Delimitação do Tema e Justificativa</b>	<b>16</b>
<b>1.2</b>	<b>Objetivos</b>	<b>17</b>
1.2.1	Objetivo Geral	17
1.2.2	Objetivos Específicos	18
<b>1.3</b>	<b>Estrutura da dissertação</b>	<b>18</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>19</b>
<b>2.1</b>	<b>Dados Abertos Governamentais</b>	<b>19</b>
2.1.1	Parceria para Governo Aberto	19
2.1.2	Impacto nas políticas públicas	20
<b>2.2</b>	<b>Legislação aplicada à Política de Dados Abertos</b>	<b>21</b>
2.2.1	Lei de Acesso à Informação	21
2.2.2	Política de Dados Abertos do Poder Executivo Federal	23
2.2.3	Lei Geral de Proteção de Dados Pessoais	24
2.2.4	Licenças Abertas	26
<b>2.3</b>	<b>Plano de Dados Abertos da UFRPE</b>	<b>28</b>
<b>2.4</b>	<b>Plataformas de gerenciamento de dados abertos</b>	<b>29</b>
<b>3</b>	<b>TRABALHOS RELACIONADOS</b>	<b>31</b>
<b>4</b>	<b>METODOLOGIA</b>	<b>36</b>
<b>4.1</b>	<b>Metodologia de pesquisa</b>	<b>36</b>
<b>4.2</b>	<b>Tecnologias adotadas</b>	<b>37</b>
<b>4.3</b>	<b>Atendimento às normas e regras</b>	<b>38</b>
<b>5</b>	<b>IMPLEMENTAÇÃO</b>	<b>40</b>
<b>5.1</b>	<b>Escolha do CKAN para gerenciamento de dados abertos</b>	<b>40</b>
5.1.1	Estudo da Literatura	40
5.1.2	Estudo da Legislação	41
<b>5.2</b>	<b>Instalação e configuração do CKAN</b>	<b>43</b>
5.2.1	Instalação	43
5.2.2	Configuração Inicial	43
5.2.3	Configurações avançadas	45

5.2.4	Instalação e Configuração de <i>Plugins</i> . . . . .	47
5.2.5	Integração com o Portal Brasileiro de Dados Abertos . . . . .	50
<b>5.3</b>	<b>Segurança</b> . . . . .	<b>52</b>
<b>5.4</b>	<b>Gerenciamento de permissões no CKAN</b> . . . . .	<b>53</b>
<b>5.5</b>	<b>Orientações para implementação do portal de dados abertos</b> . . . . .	<b>54</b>
<b>6</b>	<b>ESTUDO DE CASO</b> . . . . .	<b>55</b>
<b>6.1</b>	<b>Caso I - Ensino de Graduação</b> . . . . .	<b>55</b>
6.1.1	Dados do sistema acadêmico . . . . .	56
6.1.2	Dados do censo de educação superior . . . . .	61
<b>6.2</b>	<b>Caso II - Orçamento</b> . . . . .	<b>63</b>
<b>6.3</b>	<b>Caso III - Contratos</b> . . . . .	<b>64</b>
<b>6.4</b>	<b>Principais descobertas</b> . . . . .	<b>66</b>
<b>7</b>	<b>BOAS PRÁTICAS PARA PUBLICAÇÃO DE DADOS ABERTOS</b> . . . . .	<b>68</b>
<b>7.1</b>	<b>Extração e preparação do conteúdo</b> . . . . .	<b>68</b>
<b>7.2</b>	<b>Inclusão dos Metadados</b> . . . . .	<b>69</b>
<b>7.3</b>	<b>Processo de publicação de dados abertos</b> . . . . .	<b>71</b>
<b>7.4</b>	<b>Papéis e responsabilidades</b> . . . . .	<b>72</b>
<b>7.5</b>	<b>Inserção dos conjuntos de dados no portal</b> . . . . .	<b>73</b>
<b>8</b>	<b>CONCLUSÕES</b> . . . . .	<b>76</b>
<b>8.1</b>	<b>Contribuições e Análises dos Objetivos Propostos</b> . . . . .	<b>77</b>
<b>8.2</b>	<b>Limitações da Pesquisa e Trabalhos Futuros</b> . . . . .	<b>78</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>79</b>
	<b>APÊNDICE A – SCRIPTS DE CONFIGURAÇÃO DO CKAN</b> . . . . .	<b>83</b>

## 1 INTRODUÇÃO

A Parceria para Governo Aberto, ou *Open Government Data (OGD)*, iniciada em 2011, impulsionou a criação de políticas de transparência e abertura de dados em diversos países. O acesso aos dados governamentais permite aos cidadãos entenderem o funcionamento de seus governos e avaliarem sua eficiência, melhorando o processo de tomada de decisão e permitindo que situações de uso indevido de recursos sejam expostas e denunciadas (UBALDI, 2013).

Para Martins (2012), o direito de acesso à informação - fundamental em sociedades democráticas - impõe aos governos duas obrigações. A primeira consiste em divulgar informações essenciais acerca do funcionamento dos órgãos e instituições públicas, enquanto a segunda implica no compromisso de atender as solicitações de informação feitas por cidadãos.

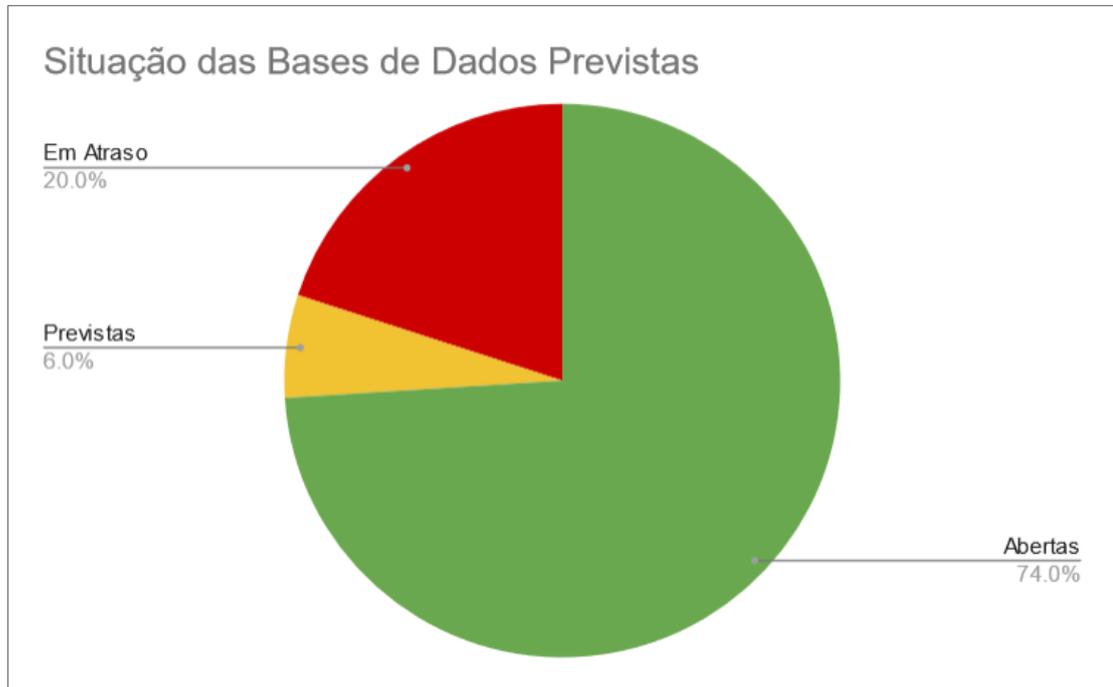
No Brasil, o acesso aos dados governamentais, previsto na Constituição Federal de 1988, ganhou força em 2012, com a publicação da Lei de Acesso à Informação (LAI), a primeira de uma série de regulamentações federais para abertura de dados. A lei prevê que os órgãos e entidades públicos deverão promover a disponibilização dos dados de interesse da população em portal na internet e responder às solicitações de informação da sociedade civil.

Um dos meios de atender a esses requisitos é através de um portal de dados abertos, onde devem ser publicadas as bases de dados previstas para abertura no Plano de Dados Abertos (PDA) da instituição (BRASIL, 2019b). No entanto, muitas entidades encontram dificuldades em atender a este requisito. Segundo o Painel de Monitoramento de Dados Abertos<sup>1</sup>, da Controladoria-Geral da União (CGU), 57,8% delas estão sem PDA, 41,4% possuem PDA publicado e disponível em seu portal, e 0,8% estão com PDA em construção. Dentre os órgãos com PDA publicado, 20% das bases previstas estão em atraso (BRASIL, 2022b), conforme Figura 1. Esse era o caso da Universidade Federal Rural de Pernambuco (UFRPE), que publicou seu PDA em 2022 e ainda não possuía um portal de dados abertos funcional.

---

<sup>1</sup> Disponível em: <http://paineis.cgu.gov.br/dadosabertos/index.htm>. Acessado em: 26 dez. 2022.

Figura 1 – Painel de Monitoramento de Dados Abertos - Bases Previstas



Fonte: Adaptado de BRASIL (2022b)

### 1.1 Delimitação do Tema e Justificativa

Apesar de fazer mais de uma década desde a publicação da LAI e da crescente cobrança por parte do Governo Federal, através da CGU, os dados do Painel de Monitoramento deixam claro que muitos órgãos e instituições encontram barreiras ao colocar em prática sua política de dados abertos.

Para Craveiro e Martano (2014), as atividades necessárias para alcançar o sucesso no processo de abertura de dados e seu impacto ainda são pouco conhecidas. A falta de uma estratégia para abertura limita a qualidade das informações e, conseqüentemente, seu potencial de reuso e geração de valor para a sociedade.

A disponibilização de um portal de dados abertos não é o bastante para atender os requisitos de transparência. Para que esta obtenha sucesso, é necessário que as informações possam ser utilizadas de fato. Ou seja, os dados precisam ser de qualidade: acessíveis, compreensíveis e facilmente reutilizáveis (REIS; MONTEIRO; SILVA, 2020).

Afinal, os benefícios dos dados abertos não estão estritamente na informação em si, mas em seu potencial de reuso. Assim, a prioridade dos órgãos não deve ser apenas tornar suas informações disponíveis, mas sim fornecer meios para que a população consiga encontrar,

visualizar e analisar esses dados (JANSSEN; CHARALABIDIS; ZUIDERWIJK, 2012).

Muitos autores relatam os desafios encontrados pelas instituições públicas em abrir os seus dados. Entre eles, pode-se mencionar o desconhecimento da legislação (CORREA; CORREA; SILVA, 2014) e dificuldade de implementar os requisitos técnicos no portal (LIMA; ABDALLA; OLIVEIRA, 2020). Como resultado, a qualidade das informações divulgadas muitas vezes não possibilita sua utilização pela sociedade (BACHTIAR; SUHARDI; MUHAMAD, 2020).

Neste contexto, está a UFRPE, uma das 69 Instituições Federais de Ensino Superior (IFES) ligadas ao Ministério da Educação (MEC). Ela oferece 55 cursos de Graduação, 56 cursos de Pós-Graduação *Stricto Sensu*, sete Técnicos e um centro de educação infantil (PDA-UFRPE, 2022). Por ser uma Autarquia Federal de Ensino, a instituição tem o dever de divulgar respostas para as perguntas mais frequentes da sociedade, assim como dados de seus programas e ações, estrutura organizacional, registros de despesas, repasses financeiros e licitações (BRASIL, 2011).

Assim, optou-se por fazer um estudo de caso com a implantação do portal de dados abertos da UFRPE. Além do fato de o desenvolvimento do portal ser uma demanda da instituição, dada a publicação recente de seu PDA, contribuiu para essa escolha o fato de a autora ser servidora da UFRPE, lotada na Secretaria de Tecnologias Digitais (STD). O estudo terá como objetivo avaliar as principais tecnologias disponíveis para publicação de dados abertos governamentais e elaborar um guia de boas práticas a ser seguido pelas IFES.

## **1.2 Objetivos**

Esta pesquisa se propõe a investigar o processo de implementação de um portal de dados abertos, comparando as aplicações disponíveis para gerenciamento de dados, e propor uma estratégia de publicação das bases previstas no PDA, com foco nas IFES, de acordo com as melhores práticas de OGD e os requisitos da legislação brasileira. Esta Seção detalha os objetivos geral e específicos a serem alcançados no presente trabalho.

### **1.2.1 Objetivo Geral**

O objetivo geral do estudo é implementar um portal de dados abertos e desenvolver um modelo de boas práticas para o tratamento e divulgação dos dados das IFES, como forma de atender aos requisitos de transparência ativa, em conformidade com a LAI.

### 1.2.2 Objetivos Específicos

De forma a atingir o objetivo geral proposto acima, foram delimitados os seguintes objetivos específicos:

1. Conduzir estudo da literatura e da legislação brasileira sobre publicação de dados abertos governamentais;
2. Realizar implementação de portal de dados abertos, atendendo aos requisitos da legislação, da Infraestrutura Nacional de Dados Abertos (INDA) e de segurança da informação.
3. Promover a integração com o Portal Brasileiro de Dados Abertos.
4. Realizar estudo de caso com três bases de dados elencadas no PDA da UFRPE, retratando desde o processo de descoberta dos dados à publicação no portal.

### 1.3 Estrutura da dissertação

Esta dissertação está organizada da seguinte forma: o Capítulo 1 define o tema, expõe a problemática e objetivos do trabalho; o Capítulo 2 apresenta a revisão da literatura para melhor contextualização do tema; o Capítulo 3 faz uma análise dos trabalhos relacionados; o Capítulo 4 contém a metodologia utilizada; o Capítulo 5 relata o estudo das tecnologias e o processo de implantação do portal; o Capítulo 6 traz um estudo de caso com bases de dados da UFRPE; o Capítulo 7 traz um guia de boas práticas para publicação de dados abertos em IFES e o Capítulo 8 mostra as conclusões da autora e recomendações de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este Capítulo apresenta os conceitos necessários para a fundamentação teórica da dissertação. Ele está estruturado da seguinte forma: a Seção 2.1 define o conceito de dados abertos governamentais e seu impacto nas políticas públicas. A Seção 2.2 detalha a legislação que rege a abertura de dados no Brasil. A Seção 2.3 apresenta o conteúdo do PDA da UFRPE e a Seção 2.4 introduz as plataformas para gerenciamento de portais de dados abertos.

### 2.1 Dados Abertos Governamentais

O conhecimento pode ser considerado aberto quando qualquer pessoa está livre para acessá-lo, utilizá-lo, modificá-lo, e compartilhá-lo, sujeito apenas à medidas que preservam a proveniência e abertura. Tal definição foi feita pela Open Knowledge Foundation (OKFN), organização que atua na promoção do conhecimento aberto e defende o direito de acessar, reutilizar e redistribuir informações (MOLLOY, 2011).

Nas três leis dos dados abertos conceitualizadas por Eaves (2009), os cidadãos precisam conseguir encontrar, utilizar e compartilhar um dado para que ele possa ser considerado aberto. Segundo o autor, para ser encontrada na internet, a informação deve estar disponível em ferramentas de busca como o Google. Para ser utilizada, ela precisa ser publicada em formato não proprietário e legível por máquina. Por fim, para o dado ser compartilhado o autor deve fornecer mecanismos legais para isso, ou seja, a informação deve ser domínio público ou fornecida sob uma licença aberta.

#### 2.1.1 Parceria para Governo Aberto

A publicação de dados abertos governamentais é regida por oito princípios, definidos em 2007 por 30 especialistas em encontro realizado na Califórnia, Estados Unidos (UBALDI, 2013). São eles:

1. Completos. Os dados devem ser divulgados integralmente, exceto quando sujeitos a limitações de privacidade ou sigilo.
2. Primários. Os dados são publicados com seu conteúdo original, com a mais fina granularidade possível.

3. Atuais. Os dados são divulgados assim que possível, de forma a preservar o seu valor.
4. Acessíveis. Os dados são disponibilizados para o público mais amplo possível para qualquer propósito.
5. Processáveis por máquina. Os dados são estruturados para possibilitar o seu processamento automatizado.
6. Não discriminatório. Não deve ser exigida identificação ou registro para liberar acesso aos dados.
7. Formatos não proprietários. Os dados devem estar em formato livre de qualquer restrição.
8. Licenças livres. Os dados devem ser publicados em licenças abertas.

Foi deste conceito que surgiu a Parceria para Governo Aberto, iniciativa internacional fundada em 20 de setembro de 2011 por oito países: África do Sul, Brasil, Estados Unidos, Filipinas, Indonésia, México, Noruega e Reino Unido. Nesta data, foi assinada a Declaração de Governo Aberto, documento que firma o compromisso em incentivar globalmente práticas governamentais de transparência fiscal, acesso à informação, participação cidadã e divulgação de declarações patrimoniais por autoridades (SHINTAKU; SALES, 2019).

A Parceria promove a abertura dos dados dos governos seguindo os princípios de transparência, responsabilidade, participação social e inovação e tecnologia. Os países membros assumem um compromisso de garantir transparência fiscal, acesso à informação, participação cidadã e divulgação de declarações patrimoniais das autoridades (SHINTAKU; SALES, 2019).

### 2.1.2 Impacto nas políticas públicas

Para Ubaldi (2013), dados governamentais são um recurso estratégico, com potencial de beneficiar diversos setores da sociedade. Esses setores incluem o próprio governo, melhorando a eficiência na alocação de recursos e reduzindo fraudes; e os cidadãos, que passam a ter meios de monitorar e responsabilizar o governo, possivelmente identificando casos de corrupção e uso indevido dos recursos públicos.

A autora ressalta a importância da participação dos funcionários públicos nas políticas de abertura de dados. Por estarem em uma posição única, como usuários e fornecedores dessas informações, esses profissionais, quando devidamente incentivados e qualificados, podem melhorar a qualidade e experiência de uso dos dados públicos (UBALDI, 2013).

As informações governamentais também são relevantes do ponto de vista econômico. Ao fornecer informações públicas sem custo, o governo promove a criação de produtos e serviços.

Isso estimula a economia e gera recursos na forma de impostos (UBALDI, 2013). A Figura 2 ilustra o processo de geração de valor a partir de dados abertos governamentais.

Figura 2 – Geração de valor a partir de dados abertos governamentais



**Fonte:** Adaptado de UBALDI (2013)

## 2.2 Legislação aplicada à Política de Dados Abertos

Esta Seção detalha os principais mecanismos da legislação brasileira que regem o processo de abertura de dados, com ênfase em IFES, e o papel das licenças abertas na publicação de dados abertos governamentais.

### 2.2.1 Lei de Acesso à Informação

O acesso à informação é um direito fundamental, garantido no Brasil pelos artigos 5, 37 e 216 da Constituição Federal de 1988 (BRASIL, 1988). O direito de acesso público à informação também é um dos pontos centrais de dois tratados internacionais ratificados pelo Estado brasileiro: o Pacto Internacional dos Direitos Civis e Políticos e a Convenção Americana de

Direitos Humanos (MARTINS, 2012).

Todos têm direito a receber dos órgãos públicos informações de seu interesse particular, ou de interesse coletivo ou geral, que serão prestadas no prazo da lei, sob pena de responsabilidade, ressalvadas aquelas cujo sigilo seja imprescindível à segurança da sociedade e do Estado. (BRASIL, 1988).

Nesse contexto, entra a Lei nº 12.527, de 2011, conhecida como LAI. Influenciada pelos princípios de dados abertos governamentais, ela regulamenta o direito de acesso dos cidadãos às informações públicas previsto na Constituição Federal e se aplica para os poderes executivo, legislativo e judiciário da União, Estados, Distrito Federal e Municípios (BRASIL, 2011).

O ponto fundamental da LAI institui o acesso à informação pública como a regra, e o sigilo como exceção. Ao longo de seus 47 artigos, a LAI fornece mecanismos para garantir a transparência dos dados de forma ativa, que se dá quando a entidade disponibiliza por conta própria as informações, e passiva, na forma de resposta aos pedidos de informação que podem ser feitos por qualquer cidadão (BRASIL, 2011).

No caso da transparência ativa, a lei determina que os órgãos e entidades públicas devem promover, independente de requerimentos, a divulgação de informações de interesse geral em portal oficial na internet. Esse portal deverá conter dados para acompanhamento das ações e programas desenvolvidos pelo órgão ou entidade, despesas e demais registros financeiros, e respostas para as perguntas mais frequentes dos cidadãos (BRASIL, 2011). A publicação de dados em formato aberto é uma dentre as várias iniciativas de transparência ativa.

A transparência passiva, por sua vez, consiste no atendimento de pedidos de informação feitos pela sociedade. Qualquer cidadão pode solicitar informações de órgãos e entidades públicas por qualquer meio legítimo - como formulário online, telefone ou pessoalmente - e tais dados devem ser disponibilizados dentro do prazo estabelecido pela lei, a menos que estejam protegidos pelos casos de sigilo previstos (BRASIL, 2011).

Comumente, tais pedidos são feitos através da Plataforma Integrada de Ouvidoria e Acesso à Informação (Fala.BR)<sup>1</sup>, desenvolvida e mantida pela CGU que, desde 2020, sucede o Sistema Eletrônico do Serviço de Informações ao Cidadão (e-SIC).

Em um artigo elaborado por Silva e Bruni (2019), a transparência ativa é descrita como sendo uma iniciativa do gestor público, que tem o dever de publicar no mínimo as informações previstas pela legislação, enquanto na transparência passiva a iniciativa é do cidadão, que tem o direito de solicitar acesso a qualquer tipo de dado. A divulgação de tal dado será realizada

<sup>1</sup> Disponível em: <https://falabr.cgu.gov.br>. Acessado em: 19 out. 2022.

desde que este não se enquadre em nenhuma das situações de sigilo previstas em lei. O Quadro 1 mostra as principais diferenças entre transparência ativa e passiva.

Quadro 1 – Diferenças entre transparência ativa e passiva

<b>Característica</b>	<b>Transparência Ativa</b>	<b>Transparência Passiva</b>
Iniciativa	Gestor público	Qualquer interessado
Forma de Acesso	Internet	Fala.BR
Conteúdo	A critério do gestor público, desde que divulgado o mínimo previsto em lei e resguardado o sigilo de informações, quando cabível.	Qualquer informação solicitada, resguardado o sigilo de informações, quando cabível.
Prazo	Tempo real	Imediatamente, quando disponível, ou no máximo 20 dias, prorrogáveis por mais 10 dias.

**Fonte:** Adaptado de SILVA; BRUNI (2019)

## 2.2.2 Política de Dados Abertos do Poder Executivo Federal

A Política de Dados Abertos do Poder Executivo Federal, regida pelo Decreto nº 8.777, de 2016 (BRASIL, 2016), alterado pelo Decreto nº 9.903, de 2019 (BRASIL, 2019a), e pela Resolução nº 3 do Comitê Gestor da INDA (CGINDA) (CGINDA, 2017b), estabelece a abertura dos dados das bases dos órgãos e entidades da administração pública federal, como dados abertos e em formato aberto. Os Decretos também determinam que a CGU é responsável por monitorar a aplicação da política de Dados Abertos na esfera federal, por meio da INDA.

Entre os principais objetivos da Política estão: promover a publicação de dados contidos em bases de dados de órgãos e entidades da administração pública federal, aprimorar a cultura de transparência pública, e permitir ao cidadão acesso aos dados produzidos ou acumulados pelo Poder Executivo Federal (BRASIL, 2016).

O PDA é o principal instrumento da Política de Dados Abertos e organiza as ações de implementação e promoção da abertura de dados dos órgãos (BRASIL, 2016). O PDA deve conter uma contextualização do órgão ou entidade e uma descrição das bases de dados que serão disponibilizadas para transparência ativa, além da justificativa para a escolha, levando em consideração seu grau de relevância para a sociedade e os dados mais solicitados em transparência passiva (CGINDA, 2017b).

A disponibilização dos dados definidos no PDA deve ser feita em uma página acessível no primeiro nível do portal do órgão através de um menu denominado “Acesso à Informação”. O portal deve conter ferramenta de busca de conteúdo, dados atualizados e informação em linguagem cidadã, de forma que seu entendimento não seja comprometido por falta de conhecimento dos termos técnicos. Estas e outras diretrizes estão no Guia de Transparência Ativa (GTA) (BRASIL, 2019b), elaborado pela CGU em consonância com a LAI, e deve ser observado por todas as entidades da esfera federal. O guia é atualizado periodicamente, sendo a versão mais recente de 2019 (LIMA; ABDALLA; OLIVEIRA, 2020).

Os dados disponibilizados pelos órgãos e unidades da esfera federal também devem estar de acordo com os Padrões de Interoperabilidade de Governo Eletrônico (e-PING), institucionalizados na Portaria Normativa nº 5, de 14 de julho de 2005, que definem um conjunto de especificações e premissas para a utilização de Tecnologia de Informação e Comunicação (TIC) em sites institucionais (BRASIL, 2005b). Também é preciso atender às recomendações do Modelo de Acessibilidade em Governo Eletrônico (e-MAG), guia desenvolvido com a proposta de guiar a criação e adaptação dos conteúdos digitais do governo federal para pessoas com deficiência, de forma a promover a inclusão social (BRASIL, 2005a).

A publicação de dados geoespaciais devem seguir os padrões da Infraestrutura Nacional de Dados Espaciais (INDE), definida pelo Decreto nº 6.666 de 2008. Nos metadados é preciso constar os Assuntos Relacionados do Vocabulário Controlado do Governo Eletrônico (VCGE) (BRASIL, 2019b), que consiste em um catálogo das informações mantidas pelo governo federal.

### 2.2.3 Lei Geral de Proteção de Dados Pessoais

A LAI considera informação pessoal aquela relacionada à pessoa natural identificada ou identificável. A Lei Geral de Proteção de Dados Pessoais (LGPD) - Lei nº 13.709, de 2018 - mantém esta definição e adiciona o conceito de dado pessoal sensível (BRASIL, 2018).

Dado pessoal sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural (BRASIL, 1988).

Para a LGPD, a informação pessoal pode ter natureza objetiva (ex. idade) ou subjetiva (ex. o devedor X é confiável), e se relaciona a um conteúdo sobre o indivíduo, com a finalidade de avaliar a ele ou seu comportamento; ou com impacto sobre seus interesses ou direitos. Para

ser pessoal, a informação deve estar relacionada a um indivíduo humano.

O termo “*Identificada*” indica uma ligação direta ao indivíduo, como seu nome completo ou foto. Já o termo “*identificável*” se aplica para uma ligação indireta, que pode exigir um processo de cruzamento de dados para a identificação. É o caso de identificadores como o Registro Geral (RG), Cadastro de Pessoa Física (CPF), o endereço e o telefone, assim como dados de localização do indivíduo.

Estas informações consideradas sensíveis possuem proteção especial, de forma a garantir o direito de autodeterminação de imagem do titular e protegê-lo de ações discriminatórias por parte de terceiros. Tal proteção já estava presente na Constituição Federal e na LAI. A LGPD reforça esse aspecto e regulamenta a forma de realizar o tratamento destes dados no contexto de transparência e acesso à informação (BRASIL, 2020a).

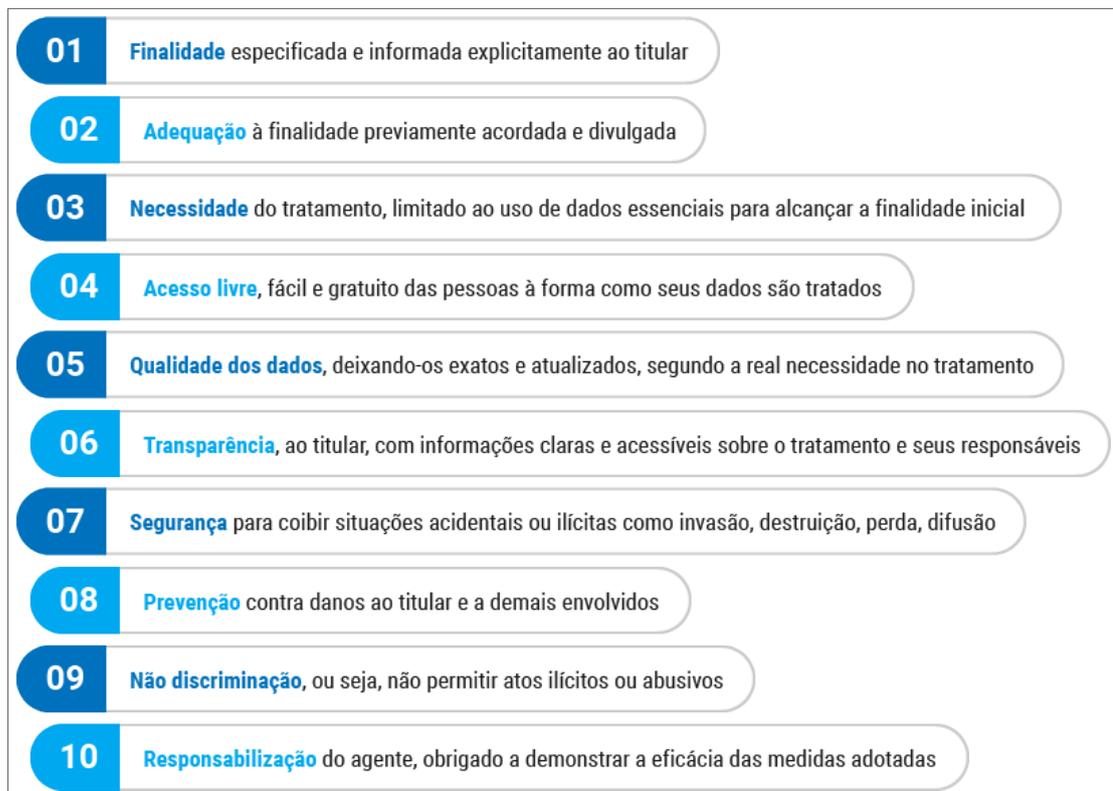
A LGPD tem muito em comum com o Regulamento Geral sobre Proteção de Dados (GDPR), a regulamentação da União Europeia para dados pessoais, também de 2018. Tanto a lei brasileira quanto a europeia têm como pilar a proteção, transparência no tratamento dos dados e direito de escolha a cerca de sua divulgação e publicidade (LORENZON, 2021).

Em seu artigo 7º, a LGPD delibera que a divulgação de informação pessoal dispensa consentimento do titular para o cumprimento de obrigação legal ou regulatória pelo controlador, no qual se enquadra a LAI. Por sua vez, o artigo 11º estipula que o tratamento e divulgação de dados sensíveis pode ser feito apenas em situações indispensáveis, entre as quais também consta o cumprimento de obrigação legal ou regulatória (BRASIL, 2018). Isso dá aos órgãos públicos permissão para tratar dados pessoais e sensíveis no cumprimento de suas obrigações legais, dispensando a autorização do titular dos dados.

Contudo, de forma a respeitar os princípios da LAI e da LGPD, não basta aos órgãos públicos enquadrar a divulgação de dados pessoais e sensíveis como cumprimento de obrigação legal. Uma das exigências definidas pela LGPD é a adesão à boa fé e aos 10 princípios fundamentais listados em seu artigo 6º (BRASIL, 2020a) e ilustrados na Figura 3.

Por fim, em seu artigo 23º, a LGPD estabelece que é responsabilidade do órgão ou entidade informar de forma pública e de fácil acesso a finalidade e hipótese legal para a utilização dessas informações pessoais. Ou seja, ao divulgar informações pessoais em cumprimento da LAI, deve ser deixado explícito para o titular como esses dados serão tornados públicos e a razão da dispensa do consentimento (BRASIL, 2020a).

Figura 3 – Princípios da LGPD



Fonte: SERPRO (2022)

#### 2.2.4 Licenças Abertas

Para serem considerados dados abertos, não basta que os conjuntos de dados, ou *datasets*, estejam em formato aberto, eles também precisam estar legalmente abertos (EAVES, 2009). Em definição da OKFN, “A obra deve obrigatoriamente estar em domínio público ou ser fornecida sob uma licença aberta”. Uma licença é considerada aberta quando permite irrevogavelmente o uso, redistribuição, modificação, separação e compilação para qualquer finalidade, sem necessidade de citar a fonte (MOLLOY, 2011).

Entre as licenças abertas disponíveis, as mais utilizadas em dados abertos governamentais são a Public Domain Dedication and License (PDDL) e Open Data Commons Open Database License (ODbL), desenvolvidas pela própria OKFN, e a Creative Commons Zero (CC0) e a Creative Commons Atribuição 4.0 (CC BY 4.0), a versão mais liberal da licença Creative Commons (CC), desenvolvida e mantida pela Organização Não Governamental (ONG) de mesmo nome localizada na Califórnia, Estados Unidos (JAAKKOLA; MAKINEN; ETELAHO, 2014).

Muitos autores questionam se existe a necessidade de licenciar dados abertos governamentais. Contudo, para Davies (2012), a licença é uma ferramenta que torna explícito tudo que

pode ser feito com determinado conjunto de dados. Portanto, o autor recomenda a utilização de licenças abertas, ou seja, sem nenhuma restrição de uso, em seu compartilhamento.

Em artigo publicado por Khayyat e Bannister (2014), o uso de uma licença aberta não só é necessário, mas essencial para o sucesso das iniciativas de dados abertos. O oferecimento de garantias legais que permitam o compartilhamento e reuso das bases de dados sem restrições encoraja o desenvolvimento de aplicações e ferramentas a partir dessas informações.

A questão do licenciamento de dados abertos governamentais é tratada de forma diferente em cada país. Governos que possuem uma política de abertura de dados mais consolidada, como Canadá, Reino Unido, Alemanha, França e Noruega, possuem uma licença própria que garante livre utilização das informações divulgadas. Já Itália e Países Baixos, por sua vez, utilizam uma versão da licença CC, com restrições variadas (KHAYYAT; BANNISTER, 2014).

No Brasil, os Decretos nº 8777 e nº 9903 estabelecem que *“Os dados disponibilizados pelo Poder Executivo federal e as informações de transparência ativa são de livre utilização pelos Poderes Públicos e pela sociedade”* (BRASIL, 2019a). É autorizada a utilização gratuita de toda informação cujo detentor de direitos autorais seja a União. A LAI, em seu Artigo 3º, fornece permissão irrestrita de reuso das bases de dados. A única ressalva é feita em relação aos dados pessoais e sensíveis, que devem receber tratamento apropriado.

Sendo assim, a princípio o uso das informações presentes nas bases de dados abertos já está protegido pela legislação brasileira. No entanto, informar explicitamente uma licença aberta fornece maior segurança para o público e deve ser adotada por instituições que desejam incentivar o uso de seus dados pela sociedade.

Por este motivo, o CGINDA determina, na Resolução nº 2, de 24 de março de 2017, que os dados pertencentes à União devem ser divulgados em uma das seguintes licenças: PDDL, ODbL, CC0 ou CC BY 4.0 (CGINDA, 2017a). O Quadro 2 ilustra um comparativo entre elas. O uso irrestrito se refere à utilização poder ter qualquer fim, comercial ou não.

A atribuição diz respeito à necessidade de atribuir crédito ao criador do conteúdo. Para Davies (2012), licenças abertas com algum tipo de restrição, como atribuição, podem dificultar a combinação dos dados com outro conjunto. Por esse motivo, o autor recomenda o uso de licenças sem atribuição na publicação de dados abertos.

Por fim, uma licença é considerada ideal para *datasets* quando tal uso é previsto em seu texto. Tal detalhe é relevante pois muitas das licenças abertas disponíveis se aplicam para propriedade intelectual. No entanto, não existe um consenso se conjuntos de dados abertos se enquadram nessa definição. Por isso, licenças como a PDDL e a CC0 deixam explícito a

adequação para tal uso (KHAYYAT; BANNISTER, 2014).

Quadro 2 – Comparativo de licenças abertas

<b>Licença</b>	<b>Uso irrestrito?</b>	<b>Exige atribuição?</b>	<b>Ideal para <i>datasets</i>?</b>
PDDL	Sim	Não	Sim
ODbL	Sim	Sim	Não
CC0	Sim	Não	Sim
CC BY 4.0	Sim	Sim	Não

**Fonte:** Autoria Própria (2023)

### 2.3 Plano de Dados Abertos da UFRPE

O PDA da UFRPE foi aprovado pelo Conselho Universitário (CONSU) em outubro de 2022. O documento é fruto do trabalho do Comitê de Transparência e Dados Abertos (CTDA) e tem vigência até 2024. A publicação, que segue as determinações da LAI e da INDA, tem o seguinte objetivo:

Promover a abertura de dados da Universidade Federal Rural de Pernambuco, zelando pelos princípios da publicidade, transparência e eficiência, visando ao aumento da disseminação de dados e informações de qualidade para a sociedade, de forma a contribuir com a tomada de decisão pelos gestores públicos e para o controle social, considerando as normas institucionais de segurança da informação. (PDA-UFRPE, 2022).

A elaboração do PDA foi fortemente baseada no trabalho de Arruda (2019), que fez uma pesquisa com a comunidade acadêmica e elaborou uma proposta de priorização e classificação dos dados que seriam abertos. O resultado da consulta pública, juntamente com as informações de dados solicitados pelos serviços de transparência passiva (na época, o e-SIC) e às determinações da CGINDA, deram origem à matriz de priorização mostrada de forma resumida no Quadro 3. A matriz define a ordem e conteúdo a ser divulgado.

Cada base de dados possui um órgão e um servidor público responsável por sua abertura, e uma frequência de atualização, que pode ser mensal, semestral ou anual. Tais servidores deverão desenvolver rotinas para a publicação dos dados extraídos dos sistemas institucionais da UFRPE e do Governo Federal sob suas responsabilidades (PDA-UFRPE, 2022).

O CTDA é responsável pelo gerenciamento das atividades previstas no PDA, que incluem o acompanhamento das informações disponibilizadas no portal pelos editores, validar se elas estão de acordo com os padrões definidos e sugerir mudanças e melhorias. O plano também

Quadro 3 – Matriz de priorização do PDA da UFRPE

Ordem	Base de dados	Descrição das bases
1	Financeiro	Dados sobre a Execução Financeira.
2	Planejamento e Gestão	Informações sobre a universidade relacionadas à administração.
3	Ensino de Graduação	Dados dos cursos de graduação presencial e a distância, Censo da Educação Superior.
4	Cargos e Funções	Lista de Funções Gratificadas e Cargos de Direção, Lista de Cargos Vagos e Ocupados.
5	Licitações	Lista de Licitações, Editais de Licitações
6	Extensão	Produtos de Extensão, Projetos de Extensão, Cursos de Extensão, Empresas vinculadas à Extensão, Bolsas de Extensão, Processos seletivos de Extensão.
7	Pós-Graduação	Dados dos cursos de pós-graduação presencial e a distância.
8	Pessoas	Lista dos Servidores Técnicos Administrativo, Lista dos Servidores Docentes, Lista dos Estagiários, Lista de Servidores Aposentados, Lista de Funcionários terceirizados, Lista de Afastamentos, Lista de Progressões, Lista de Capacitações.
9	Contratos	Lista de Contratos, Contratos vigentes.
10	Orçamento	Dados sobre a Execução orçamentária.

Fonte: Adaptado de PDA-UFRPE (2022)

define que a STD deverá hospedar o portal de dados abertos e fornecer os dados dos sistemas institucionais em formato legível por máquina (PDA-UFRPE, 2022).

Seguindo as recomendações da INDA, o documento define que todas as bases de dados disponibilizadas no portal da instituição precisam estar catalogadas também no Portal Brasileiro de Dados Abertos<sup>2</sup>, preferencialmente de forma automatizada. Os conjuntos de dados têm que conter no mínimo as seguintes informações: nome, descrição, etiquetas, assuntos do VCGE, nome e e-mail do setor responsável, periodicidade de atualização e escopo temporal (como diário, semanal, semestral ou anual, por exemplo) (PDA-UFRPE, 2022).

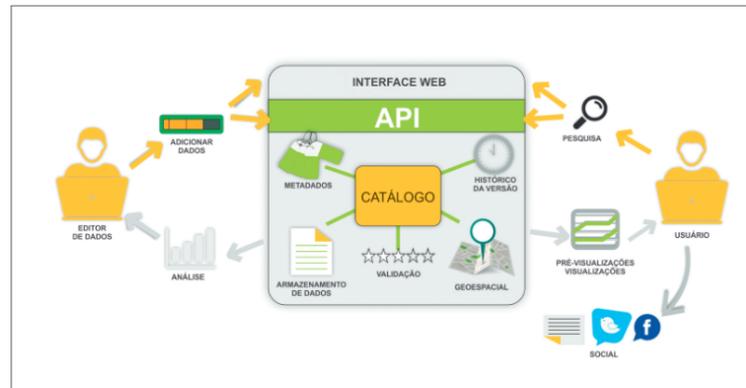
## 2.4 Plataformas de gerenciamento de dados abertos

Uma plataforma de gerenciamento de dados abertos, ou Open Data Platform (ODP), consiste em uma solução de software integrada composta por um portal, um gerenciador de metadados, biblioteca de Interface de Programação de Aplicação (API), serviços de busca e descoberta, ferramentas de visualização e relatórios. A ferramenta deve permitir publicação,

<sup>2</sup> Disponível em: <https://dados.gov.br>. Acessado em: 23 fev. 2023.

compartilhamento e visualização de dados, servindo como ponto central onde o público pode descobrir, buscar, analisar e baixar as informações de determinada organização (OSAGIE et al., 2015 apud BRASIL, 2020b).

Figura 4 – Fluxo de informação



Fonte: COSTA et al. (2017)

Segundo o modelo ilustrado na Figura 4, a aplicação fornece uma interface para World Wide Web (WEB) que permite aos responsáveis depositar conjuntos de dados diretamente ou através de uma API. Estes são armazenados e validados, podendo receber metadados, identificadores para controle de versão e localizadores, denominados tags ou etiquetas, que facilitam sua recuperação pelos usuários. Os editores podem adicionar e analisar os conjuntos de dados inseridos, enquanto os usuários podem pesquisar, visualizar e compartilhar as informações disponibilizadas (COSTA et al., 2017).

Diversas ODPs são disponibilizadas no modelo *Open Source*. Entre elas, se destacam o Dataverse<sup>3</sup>, utilizado como repositório para pesquisa científica; o DSpace<sup>4</sup>, também voltado para publicações acadêmicas; o Invenio<sup>5</sup>, utilizado como repositório institucional; e o Comprehensive Knowledge Archive Network (CKAN)<sup>6</sup>, destinado à publicação de dados abertos governamentais (CAMPELO; NETO, 2020).

<sup>3</sup> Disponível em: <https://dataverse.org>. Acessado em: 10 abr. 2023.

<sup>4</sup> Disponível em: <https://dspace.lyrasis.org>. Acessado em: 10 abr. 2023.

<sup>5</sup> Disponível em: <https://inveniosoftware.org>. Acessado em: 10 abr. 2023.

<sup>6</sup> Disponível em: <https://ckan.org>. Acessado em: 10 abr. 2023.

### 3 TRABALHOS RELACIONADOS

Este Capítulo apresenta trabalhos relacionados ao tema de dados abertos governamentais e faz uma análise comparativa com o objetivo desta pesquisa. Em *“Dados Abertos Governamentais: Uma Proposta de Classificação e Estruturação para Abertura dos Dados em Instituições Federais de Ensino Superior”*, elaborado por Arruda (2019), o autor propõe um modelo para classificação e estruturação das bases de dados que as IFES podem utilizar no processo de abertura de dados, de forma a atender aos requisitos de transparência ativa da LAI.

Com esse objetivo, foi feito um estudo das informações disponibilizadas no Portal Brasileiro de Dados Abertos, no Portal de Monitoramento de Dados Abertos e nos portais de dados abertos de todas as 63 IFES existentes na data da realização da pesquisa. Em seguida, o autor apresentou uma proposta de priorização das bases a serem abertas, levada ao CTDA da UFRPE. Essa proposta, validada em grupo focal composto por cinco especialistas e pesquisadores na área de dados abertos, foi fundamental para a elaboração do PDA da instituição.

O modelo descrito pelo autor serve como ponto de partida para a elaboração do PDA de IFES. Já esta pesquisa se diferencia por focar na implementação do PDA, abordando desde a escolha e implantação da plataforma de gerenciamento de dados abertos à elaboração de um guia de boas práticas para publicação das informações por parte dos responsáveis, cumprindo os requisitos técnicos e negociais impostos pela legislação.

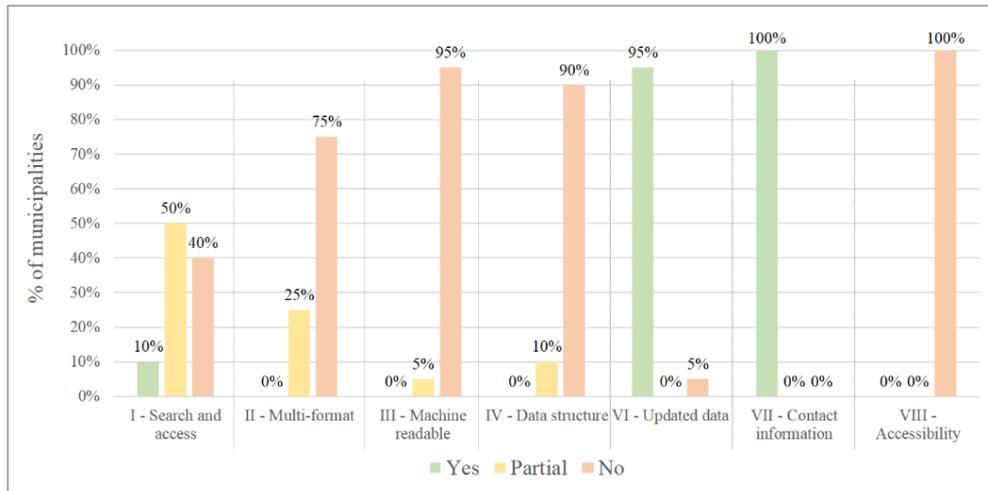
Em artigo de Correa, Correa e Silva (2014), intitulado *“Transparency Portals Versus Open Government Data. An Assessment of Openness in Brazilian Municipalities”*, os autores analisam a aderência dos portais de transparência ativa de 20 municípios brasileiros aos princípios de OGD. Eles identificaram o baixo nível de maturidade dos governos locais em relação aos requisitos da LAI como a principal causa de insucesso de sua aplicação. Segundo os autores, a maioria dos portais de transparência ativa são elaborados ignorando tanto os fundamentos de OGD como da própria LAI.

Dos critérios avaliados pelo trabalho - como uso de formatos abertos, existência de mecanismo de busca, acessibilidade, informações para contato e atualizações periódicas - 44% não foram atendidos, 32% foram atendidos parcialmente e apenas 24% estavam de acordo com a legislação. Dentre os portais que atenderam aos requisitos de forma parcial, apenas 10% utilizaram formatos abertos.

A Figura 5 mostra a adesão dos portais analisados pelos autores às boas práticas de

OGD: disponibilização de ferramenta de busca, múltiplos formatos, processável por máquina, metadados, autenticidade e integridade, dados atuais, acessibilidade e informações de contato. O quinto item, autenticidade e integridade, não foi inserido por não se enquadrar no escopo previsto para a avaliação: “Sim”, “Não” ou “Parcial”.

Figura 5 – Adesão dos portais de transparência aos princípios de OGD



Fonte: CORREA; CORREA; SILVA (2014)

O artigo também aponta a discrepância entre os portais analisados e o próprio portal nacional de dados abertos, referência internacional de uso do CKAN de acordo com o OGD, e defende a fiscalização por parte do governo federal e mudança de cultura nos governos e órgãos locais como sendo a melhor forma de mudar esse paradigma. Já este trabalho foca em analisar e demonstrar as soluções para tais desafios.

No trabalho feito por Lima, Abdalla e Oliveira (2020) - “A Avaliação da Transparência Ativa e Passiva das Universidades Públicas Federais do Brasil à Luz da Lei de Acesso à Informação” - os autores se propuseram a avaliar a aderência de 63 universidades federais brasileiras às determinações da LAI para transparência ativa e passiva. O resultado foi majoritariamente favorável no caso da transparência passiva, com 71% das instituições atendendo corretamente aos critérios de tempo de resposta e qualidade da informação divulgada. Já na avaliação da transparência ativa, esse percentual cai para 56% dos casos.

Um dos problemas identificados pelos autores nos portais de transparência ativa das universidades consiste no descumprimento das especificações do GTA. No caso dos portais de dados abertos, apenas 24 das 63 universidades analisadas atendiam aos requisitos. Os autores concluem que, mesmo com a LAI, ainda existe nessas instituições uma cultura de não prestação das informações, e aponta o desconhecimento do guia da CGU e da LAI por parte das

universidades e seus servidores como principal causa para isso. Esse mapeamento dos requisitos e boas práticas de dados abertos governamentais é uma das propostas deste trabalho.

Em *“Uma Proposta de Modelo de Processo para Publicação de Dados Abertos Conectados Governamentais”*, Avila (2015) propõe um modelo de processo, com atividades incrementais, para guiar instituições governamentais a publicarem seus dados em formatos abertos e conectados. Na questão de identificação das informações que serão disponibilizadas, o autor destaca os passos de analisar os sistemas informacionais de onde as informações serão extraídas, identificar casos em que se aplique o sigilo e medir o esforço para abertura dos dados. Já na parte de preparação dos dados para abertura, umas das prioridades deve ser sua limpeza, com identificação de possíveis erros e inconsistências.

Para o tratamento de dados sensíveis, o autor sugere três opções de anonimização para por em prática o que foi regulado na LGPD: remoção das colunas contendo a informação; agregação, divulgando apenas a média ou quantitativo de dados; ou remoção de conexões, limitando que o público cruze as informações e identifique a pessoa a quem o dado se aplica.

O autor também ressalta a importância da documentação dos dados. Isso inclui definir corretamente os nomes das colunas de dados tabulares, de forma que fique compreensível para os usuários. O uso de dicionário de dados e vocabulários também é importante para melhorar o entendimento das informações contidas no conjunto de dados. O presente trabalho aplica na prática alguns dos conceitos apresentados por Avila (2015), através de um estudo de caso com a implementação do portal de dados abertos de uma IFES.

O trabalho de Oliveira e Fonseca (2021), intitulado *“Fatores de Sucesso Na Abertura De Dados: O Caso do Banco Central do Brasil”*, analisa os fatores por trás do sucesso das políticas de abertura de dados através de um estudo de caso com o Banco Central do Brasil (BC), mostrando como a instituição superou desafios comuns à administração pública no cumprimento das diretrizes de dados abertos do governo federal.

Uma das principais conclusões da pesquisa foi o fator determinante da cultura de transparência no BC, que resultou em maior aceitação das medidas relacionadas à abertura de dados e menor resistência para que o PDA da instituição fosse bem executado.

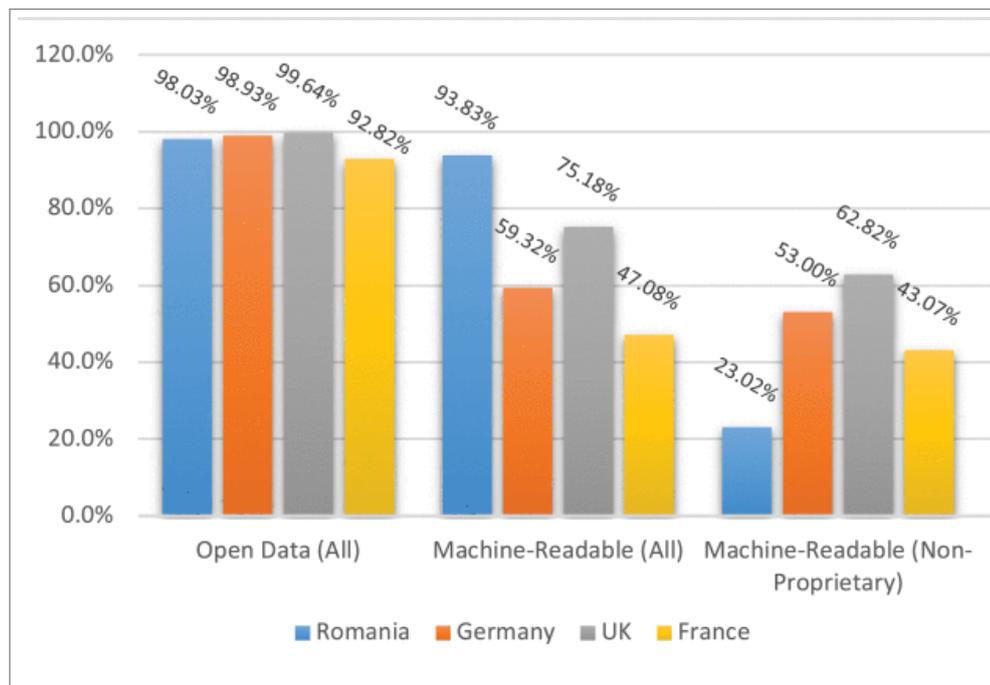
O artigo também destaca a importância do apoio da alta administração, do engajamento institucional e da participação de servidores chave no processo, citando a alta rotatividade em alguns setores da administração pública federal como um obstáculo para a abertura de dados. Os fatores de sucesso identificados pelos autores foram colocados em prática no estudo de caso realizado no presente trabalho.

O artigo de Bachtiar, Suhardi e Muhamad (2020), *“Literature Review of Open Government Data”* apresenta, através de uma revisão da literatura, os principais desafios dos governos no processo de abertura de seus dados. Em relação ao portal em si, os principais problemas encontrados pelos autores foram a disponibilidade e interoperabilidade.

Já em relação aos dados divulgados, o artigo apontou que eles muitas vezes estão em diferentes formatos dentro do mesmo portal, incorretos, incompletos, e sem informação de metadados. Outro desafio encontrado está no licenciamento dos dados como forma de incentivar e garantir o reuso e na privacidade e segurança dos dados pessoais. No entanto, o principal obstáculo encontrado pelos autores está na qualidade da informação divulgada. Uma das propostas desse trabalho consiste em explorar formas de melhorar a qualidade dos dados abertos, através de um guia de boas práticas a ser seguido pelos responsáveis pela publicação.

O artigo de Dorobăț e Posea (2021), *“Open Data Indicator: An Accumulative Methodology for Measuring the Quality of Open Government Data”* propõe um modelo para avaliar a qualidade dos dados abertos. Ao longo do estudo, foram identificados problemas em comum nas iniciativas de transparência dos governos do Reino Unido, França, Alemanha e Romênia.

Figura 6 – Qualidade dos conjuntos de dados



Fonte: DOROBĂȚ; POSEA (2021)

Conforme mostrado na Figura 6, o critério de publicação como dado aberto, ou seja, sob uma licença aberta, é atendido em quase todos os conjuntos de dados dos países analisados. No entanto, o critério de uso de formatos legíveis por máquina teve um resultado bem diferente.

A România, por exemplo, publica quase todos os seus conjuntos de dados em formato legível por máquina, mas em formatos proprietários. Já a França utiliza formatos legíveis por máquina em menos da metade das bases publicadas.

Por fim, os autores concluem que, apesar do grande volume de dados divulgados pelos países, a possibilidade de utilização é reduzida devido à falta adesão às boas práticas de publicação. Esta pesquisa tem entre seus objetivos reduzir a ocorrência deste problema.

Os trabalhos discutidos neste capítulo podem ser divididos em dois grupos. O primeiro relata desafios e dificuldades no processo de abertura de dados de instituições governamentais, enquanto o segundo traz casos de sucesso e orientações a serem seguidas. O principal diferencial desse trabalho consiste em propor uma forma de solucionar estes desafios, tendo como base as orientações de outros trabalhos e a pesquisa realizada.

## 4 METODOLOGIA

A metodologia científica é um conjunto de etapas ordenadas a serem realizadas no processo de investigação de um fenômeno. Isso inclui escolha do tema, planejamento, desenvolvimento metodológico, coleta de dados, análise, elaboração das conclusões e divulgação de resultados (SILVA; MENEZES, 2005).

Neste trabalho, são adotadas as estratégias de pesquisa bibliográfica, que consiste no estudo de material já publicado, como livros e artigos, e pesquisa documental, elaborada a partir de fontes que não receberam tratamento analítico. Tal estudo resultou no desenvolvimento de uma solução a ser validada através de um estudo de caso, que pode ser descrito como uma análise de um ou poucos objetos de maneira que se permita o seu amplo e detalhado conhecimento (SILVA; MENEZES, 2005). O presente Capítulo está estruturado da seguinte forma: a Seção 4.1 descreve a metodologia de pesquisa, a Seção 4.2 lista as tecnologias utilizadas e a Seção 4.3 relata as normas e regras a serem atendidas pelos portais de dados abertos.

### 4.1 Metodologia de pesquisa

A etapa inicial desta pesquisa consistiu em pesquisa bibliográfica. Foram analisados trabalhos relevantes para a temática de dados abertos governamentais, de forma a conceitualizar o tema a ser estudado, além de identificar as práticas realizadas no Brasil e no mundo. Esta etapa da pesquisa também possibilitou a descoberta de alguns dos problemas e desafios comuns às instituições no processo de abertura de seus dados.

Em seguida, foi feita uma pesquisa documental com a legislação vigente acerca de dados abertos e os guias publicados pelo governo. Em relação à legislação, foi preciso incluir a LAI, a Política de Dados Abertos do Governo Federal, uma vez que as IFES se enquadram na esfera federal, e a LGPD. Também serviram de fonte os manuais elaborados pela CGU.

A pesquisa bibliográfica e documental serviu como base para a etapa seguinte, que consistiu em identificar os requisitos técnicos que devem ser adotados pelos portais de dados abertos, incluindo as funcionalidades disponíveis para os usuários.

Uma vez que tais requisitos foram devidamente mapeados, foi possível escolher a aplicação de gerenciamento de dados abertos que seria utilizada na implementação do portal, assim como as alterações que seriam necessárias para atendê-los, e colocar o portal no ar.

A proposta do estudo de caso feito com a UFRPE era de selecionar três bases de dados do PDA da instituição e fazer a publicação dos dados previstos na matriz de priorização. Essa publicação foi guiada pelos requisitos negociais mapeados na etapa de pesquisa.

As bases de dados escolhidas para o estudo de caso foram: Ensino Graduação, Contratos e Orçamento. A seleção foi feita com o objetivo de estudar três diferentes casos de estratégia de publicação. Os dados da base de Ensino de Graduação precisam ser extraídos do sistema acadêmico utilizado pela instituição. Os de Contratos vêm de sistemas do Governo Federal que já possuem suporte para exportação de relatório em formato aberto. O mesmo vale para as informações de Financeiro e Orçamento, sendo que neste caso o responsável já realizava um trabalho de tratamento e análise.

O estudo resultou na publicação das bases previstas, em processo que envolveu os responsáveis por cada uma na seleção dos dados, elaboração do dicionário de dados e validação. As lições aprendidas ao longo do processo foram condensadas em um guia de boas práticas para publicação de dados abertos por IFES.

## 4.2 Tecnologias adotadas

Esta Seção apresenta as principais tecnologias utilizadas no desenvolvimento deste trabalho, incluindo softwares e linguagens de programação. Para os testes iniciais com o CKAN, a aplicação foi instalada em uma Máquina Virtual (VM). Isso permitiu testar a compatibilidade dos *plugins* e verificar seu funcionamento em um ambiente controlado. Uma vez que as configurações da ferramenta foram definidas, o processo foi replicado em produção.

Tanto o CKAN quanto o sistema acadêmico utilizado pela UFRPE utilizam o PostgreSQL como Sistema de Gerenciamento de Banco de Dados (SGBD). Para o processo de extração destes dados, se mostrou necessária instalação do módulo *tablefunc*<sup>1</sup> do PostgreSQL. Ele adiciona uma série de funções, que permitem manipular e rotacionar as linhas e colunas de uma consulta. A instalação é feita via *psql*. Além da Structured Query Language (SQL), a linguagem procedural do PostgreSQL, PL/pgSQL<sup>2</sup>, foi utilizada para criação de funções para manipulação dos dados.

Para finalizar, as linguagens de programação Python, HyperText Markup Language (HTML) e Cascading Style Sheets (CSS) foram empregadas em ajustes no código fonte do CKAN, como

<sup>1</sup> Disponível em: <https://www.postgresql.org/docs/current/tablefunc.html>. Acessado em: 18 dez. 2022.

<sup>2</sup> Disponível em: <https://www.postgresql.org/docs/current/plpgsql-overview.html>. Acessado em: 18 dez. 2022.

correções de *bugs* e melhorias de usabilidade. Tais alterações foram efetuadas na ferramenta Visual Studio Code<sup>3</sup>, editor de código fonte desenvolvido pela Microsoft como software livre.

### 4.3 Atendimento às normas e regras

Esta Seção apresenta um relato das normas e regras impostas pela legislação brasileira para publicação de dados abertos de IFES. Tais diretrizes foram utilizadas para conduzir o estudo de caso com a UFRPE e posterior elaboração de um guia de boas práticas.

Segundo diretrizes da LAI e do Decreto nº 8777, as regras negociais a serem observadas nos portais de dados abertos são: garantir reuso irrestrito das bases; designar o responsável por sua atualização e manutenção; informar a periodicidade de atualização; fornecer descrição detalhada dos *datasets* e seus campos em um dicionário de dados e utilizar linguagem de fácil compreensão para o cidadão (BRASIL, 2019b). Além de atender a esses critérios, as instituições devem se atentar ao tratamento de dados pessoais e dados pessoais sensíveis.

Em adição a essas regras, os PDAs das instituições podem trazer outras definições, detalhando alguns aspectos da implementação da LAI, que também devem ser seguidas. No caso da UFRPE, tais premissas incluem informar etiquetas ou palavras-chave, nome e e-mail do setor responsável pelos dados e periodicidade.

A LAI lista em seu artigo 8º as informações que devem constar nos portais de transparência ativa. Um dos itens consiste em “*Dados gerais para o acompanhamento de programas, ações, projetos e obras de órgãos e entidades*” (BRASIL, 2011). No caso das IFES, estas informações seriam relativas às atividades de ensino, que frequentemente possuem dados pessoais dos estudantes e docentes da instituição e precisam ser tratados seguindo as normas estabelecidas.

Por determinação da LGPD, a divulgação de dados pessoais deve ser sempre acompanhada de justificativa, com a hipótese legal que dispensa o consentimento do titular. Conforme exposto na Seção 2.2.3, São considerados dados pessoais qualquer informação pertinente à pessoa natural identificada ou identificável, como nome, data de nascimento, CPF e endereço. No caso de portais de dados abertos, todo conjunto que inclua dados pessoais deve conter a hipótese legal que permite sua publicação.

Já os dados pessoais sensíveis - onde se enquadram informações biométricas e genéticas - devem ser removidos, caso permitam a identificação do titular, ou anonimizados, exceto nos casos indispensáveis, conforme definido no Artigo 11 da LGPD. Nestas situações, a justificativa

<sup>3</sup> Disponível em: <https://code.visualstudio.com/docs>. Acessado em: 30 nov. 2022.

legal deve ser divulgada juntamente com os dados, de forma pública e acessível para o titular.

Quadro 4 – Termos do VCGE para as bases do PDA

<b>Base de dados</b>	<b>Termo</b>
Planejamento e Gestão	Planejamento e Gestão
Ensino de Graduação	Educação superior
Cargos e Funções	Recursos humanos
Licitações	Compras governamentais
Extensão	Outros em Educação
Pós-Graduação	Educação superior
Pessoas	Recursos humanos
Contratos	Compras governamentais
Orçamento e Financeiro	Orçamento

**Fonte:** Aatoria Própria (2023)

Quanto à utilização de licenças abertas, foi escolhida a PDDL. A licença atende a todos os critérios para utilização em dados abertos e se destaca devido à simplicidade de utilização, que requer apenas uma menção à licença no portal, no conjunto de dados ou dicionário de dados, e um link para a página da OKFN com sua descrição (OKFN, 2022).

Para que o portal de dados abertos das IFES atenda às recomendações expostas nessa Seção e listadas no Quadro 5, é necessário a adoção de uma série de boas práticas - ou metodologias de trabalho - por parte dos responsáveis.

Quadro 5 – Requisitos negociais para portais de dados abertos

<b>Requisito</b>	<b>Legislação</b>	<b>Como atender</b>
Informar detalhes do dataset	LAI, D8777 e PDA	Metadados
Atualizações periódicas	LAI, D8777 e PDA	Boas práticas
Instruções para contato	LAI e PDA	Campos do CKAN
Publicidade como regra	D8777	Boas práticas
Permissão irrestrita de reuso	D8777	Informar licença
Completeness dos dados	D8777	Boas práticas
Informar responsável	D8777 e PDA	Campos do CKAN
Tratar dados pessoais	LGPD	Boas práticas
Tratar dados sensíveis	LGPD	Boas práticas
Informar etiquetas	PDA	Boas práticas
Informar assunto VCGE	PDA	Boas práticas

**Fonte:** Aatoria Própria (2023)

## 5 IMPLEMENTAÇÃO

Este Capítulo traz o relato do processo de desenvolvimento e implantação de um portal de dados abertos, começando pela escolha da ODP, assim como os ajustes e configurações feitos a fim de atender às exigências da legislação brasileira, do PDA e às boas práticas de segurança da informação.

A Seção 5.1 aborda o processo de decisão acerca da aplicação para gerenciamento de dados abertos. A Seção 5.2 aborda todo o processo de instalação do CKAN. A Seção 5.3 relata os ajustes feitos na área de segurança da informação. A Seção 5.4 contém as regras para os papéis atribuídos aos usuários do sistema. Por fim, a Seção 5.5 resume o que foi aprendido ao longo do processo de implementação do portal em um guia de boas práticas. Os *scripts* de configuração utilizados podem ser encontrados no Apêndice A.

### 5.1 Escolha do CKAN para gerenciamento de dados abertos

A primeira etapa do processo de implementação do portal de dados abertos consiste na escolha da aplicação para gerenciamento de dados abertos. Com base no estudo da literatura e da legislação brasileira acerca dos requisitos a serem atendidos pelos portais, julgou-se que o CKAN era a melhor alternativa disponível no modelo *Open Source*.

#### 5.1.1 Estudo da Literatura

O estudo da literatura mostrou que o CKAN é referência mundial para portais de dados abertos governamentais. Entre os governos que utilizam a aplicação, estão: Brasil ([dados.gov.br](http://dados.gov.br)), Estados Unidos ([data.gov](http://data.gov)), Reino Unido ([data.gov.uk](http://data.gov.uk)), Canadá ([canada.ca](http://canada.ca)), Austrália ([data.gov.au](http://data.gov.au)) e Indonésia ([data.go.id](http://data.go.id)) (BACHTIAR; SUHARDI; MUHAMAD, 2020).

Em artigo elaborado por Campelo e Neto (2020), que comparou quatro softwares gratuitos para criação de portais de dados abertos (Dataverse, Invenio e Dspace), os autores demonstram que o CKAN, apesar de possuir limitações, como a falta de suporte nativo a protocolos de interoperabilidade entre repositórios, é a plataforma mais eficiente para uso em portais de dados governamentais, o que justifica sua adesão em massa pelos governos.

Desenvolvido e mantido pela OKFN, organização sem fins lucrativos fundada em 2004,

CKAN também sobressai em relação aos demais devido à sua flexibilidade e modularidade e devido ao fato de dispensar conhecimentos técnicos por parte dos responsáveis por manter as bases de dados (CAMPELO; NETO, 2020). Uma conclusão parecida foi obtida em artigo elaborado por OSAGIE et al. (2015), que compara o CKAN com DKAN<sup>1</sup>, Socrata<sup>2</sup> e Semantic MediaWik<sup>3</sup>. A aplicação foi considerada a com melhor usabilidade, além de atender nativamente ou com uso de *plugins* os principais critérios de *OGD*.

### 5.1.2 Estudo da Legislação

Ao elaborar seu portal de dados abertos, as IFES devem se atentar não só ao seu PDA, como também às normas estabelecidas pela legislação e às práticas de dados abertos governamentais. A LAI e o Decreto nº 8777 fornecem diretrizes acerca de quais dados devem ser abertos e listam requisitos técnicos a serem seguidos. Conforme será demonstrado a seguir, o CKAN se mostrou capaz de atender a todas essas exigências.

Tais requisitos técnicos podem ser resumidos em alguns pontos centrais. O primeiro deles consiste em fornecer acesso automatizado por sistemas externos em formatos abertos, estruturados e legíveis por máquina, e é atendido pelo CKAN através de seu *plugin DataStore*, que fornece uma API para leitura, busca e filtro dos conjuntos de dados presentes no portal.

Quanto aos formatos, as bases de dados devem ser publicadas exclusivamente em formatos não proprietários. O Governo Federal recomenda especificamente Comma-Separated Values (CSV), Notação de Objetos JavaScript (JSON), Extensible Markup Language (XML), Open Document Spreadsheet (ODS) e Resource Description Framework (RDF).

A próxima exigência diz respeito à permissão de gravação de relatórios em formatos abertos. Todos os conjuntos de dados presentes no CKAN podem ser baixados em seu formato original, que conforme descrito anteriormente, deve ser de um tipo não proprietário.

Outro requisito para portais de dados abertos envolve a disponibilização de ferramenta de pesquisa de conteúdo. O CKAN conta nativamente com a plataforma de busca integrada. O campo de pesquisa é um dos componentes do cabeçalho do portal, sendo exibido em todas as páginas, e pode ser configurado para aparecer em destaque na página inicial.

A legislação determina que o portal deve oferecer garantia de autenticidade e integridade das informações nele publicadas. No CKAN, apenas usuários autorizados podem criar, editar

<sup>1</sup> Disponível em: <https://getdkan.org>. Acessado em: 20 abr. 2023.

<sup>2</sup> Disponível em: <https://dev.socrata.com>. Acessado em: 20 abr. 2023.

<sup>3</sup> Disponível em: <https://www.semantic-mediawiki.org>. Acessado em: 20 abr. 2023.

e excluir os conjuntos de dados publicados. A aplicação possui campos para que os dados - como nome e endereço do e-mail - do criador e mantenedor dos *datasets* sejam informados.

Por fim, os portais de dados abertos devem:

Adotar as medidas necessárias para garantir a acessibilidade de conteúdo para pessoas com deficiência, nos termos do art. 17 da Lei nº 10.098, de 19 de dezembro de 2000, e do art. 9º da Convenção sobre os Direitos das Pessoas com Deficiência, aprovada pelo Decreto Legislativo nº 186, de 9 de julho de 2008 (BRASIL, 2011).

Apesar de não ter acessibilidade como foco, o CKAN atende parcialmente os requisitos do e-MAG. Uma das principais limitações, a falta de suporte nativo para visualização em Libras, pode ser contornada através da utilização do *VLibras Widget*<sup>4</sup>, um recurso mantido pelo Governo Federal que possibilita tradução automática de páginas da WEB para a Língua Brasileira de Sinais (LIBRAS).

Outro ponto extremamente relevante para a utilização do CKAN foi o fato de a aplicação possibilitar a sincronização automática com o Portal Brasileiro de Dados Abertos<sup>5</sup>. Lançado em dezembro de 2011, o portal do Governo Federal é uma ferramenta para catalogar as bases de dados criadas por órgãos e entidades da administração pública e serve como um ponto central para a busca de dados públicos no Brasil (BRASIL, 2019b). O Quadro 6 apresenta um resumo dos requisitos técnicos atendidos pelo CKAN, cujos detalhes de instalação e configuração são apresentados na Seção 5.2.

Quadro 6 – Requisitos técnicos para portais de dados abertos

<b>Requisito</b>	<b>Possível no CKAN</b>	<b>Tipo de configuração</b>
Acesso por API	Sim	Plugin
Formato aberto	Sim	Parâmetro
Gravação de relatórios	Sim	Nativo
Ferramenta de busca	Sim	Nativo
Autenticidade e integridade	Sim	Nativo
Atender ao e-MAG	Sim	VLibras
Sincronização automática dados.gov.br	Sim	Configuração

**Fonte:** Autoria Própria (2023)

<sup>4</sup> Disponível em: <https://vlibras.gov.br/doc/widget/index.html>. Acesso em: 22 fev. 2023.

<sup>5</sup> Disponível em: <https://dados.gov.br>. Acessado em: 20 abr. 2023.

## 5.2 Instalação e configuração do CKAN

Esta Seção aborda o processo de instalação e configuração da aplicação CKAN no portal de dados abertos da UFRPE. A Subseção 5.2.1 aborda o processo de montagem do ambiente em sua configuração padrão, seguido pelos ajustes de personalização básicos, na Subseção 5.2.2, e avançados, na Subseção 5.2.3.

A Subseção 5.2.4 descreve as extensões utilizadas e o propósito de cada uma delas. Por último, a Subseção 5.2.5 traz as configurações de organizações e grupos conforme o PDA e o processo de integração com o Portal Brasileiro de Dados Abertos.

### 5.2.1 Instalação

O CKAN foi desenvolvido majoritariamente na linguagem de programação Python. A aplicação utiliza o sistema de template Jinja e a biblioteca de JavaScript JQuery. Feita para funcionar em navegador WEB, ela utiliza o Servidor HTTP Apache, da Apache Software Foundation (ASF), como servidor WEB, e o Nginx como proxy.

O armazenamento de dados é feito no SGBD PostgreSQL. Também fazem parte dos recursos utilizados as plataformas de pesquisa Solr, da ASF, que roda no servidor Jetty e armazena seus dados no Redis, um servidor de armazenamento chave-valor (COSTA et al., 2017), conforme ilustrado na Figura 7.

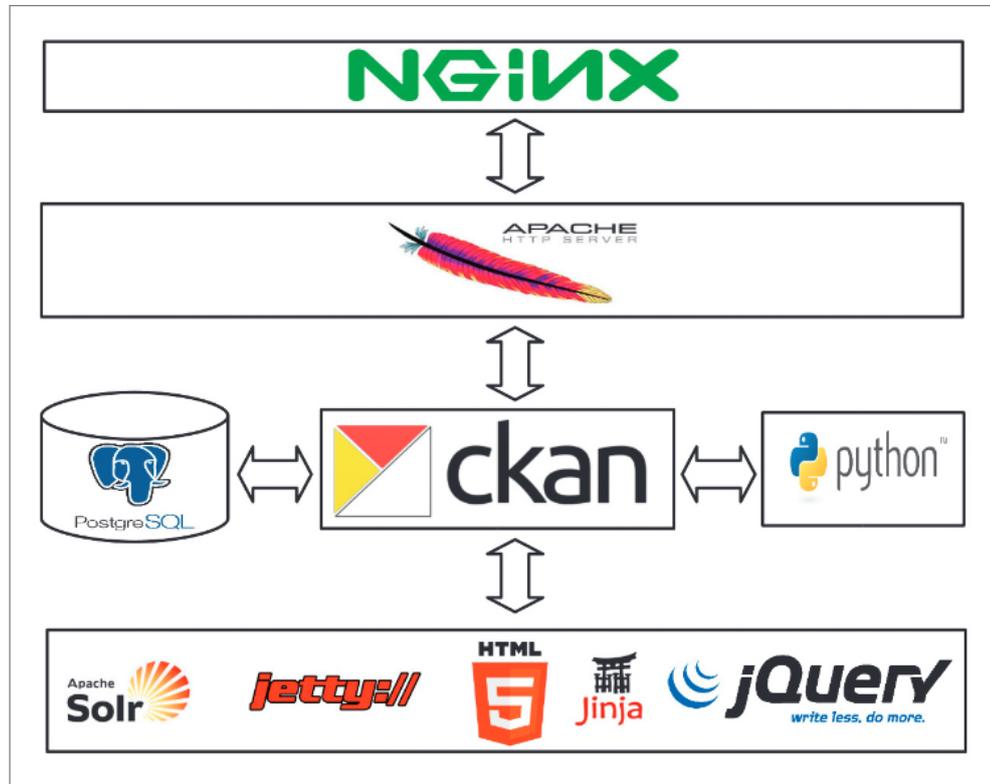
A versão corrente do CKAN - 2.9.5 - requer a versão 18.04 ou 20.04 do Ubuntu, sendo recomendado o Python 3, a partir da versão 3.6. Neste trabalho, a instalação foi feita com o Ubuntu 20.04 rodando em uma VM, com Python 3.8. O processo, que envolve a instalação do próprio CKAN, do PostgreSQL, do Solr e do plugin DataPusher - usado para gerar as visualizações dos recursos - resulta nos serviços listados no Quadro 7. Por não haver necessidade de alterar as portas, foram mantidos os valores padrão.

### 5.2.2 Configuração Inicial

Uma vez que os serviços do CKAN estejam devidamente configurados, é feita a personalização inicial da aplicação. Existem dois tipos de configuração, pela interface WEB e pelo arquivo de configuração principal. A Figura 8 retrata a aparência padrão do sistema.

Pela interface, são alterados alguns aspectos da página inicial, como o título e logo do

Figura 7 – Estrutura do CKAN



Fonte: COSTA et al. (2017)

Quadro 7 – Serviços que compõe o CKAN

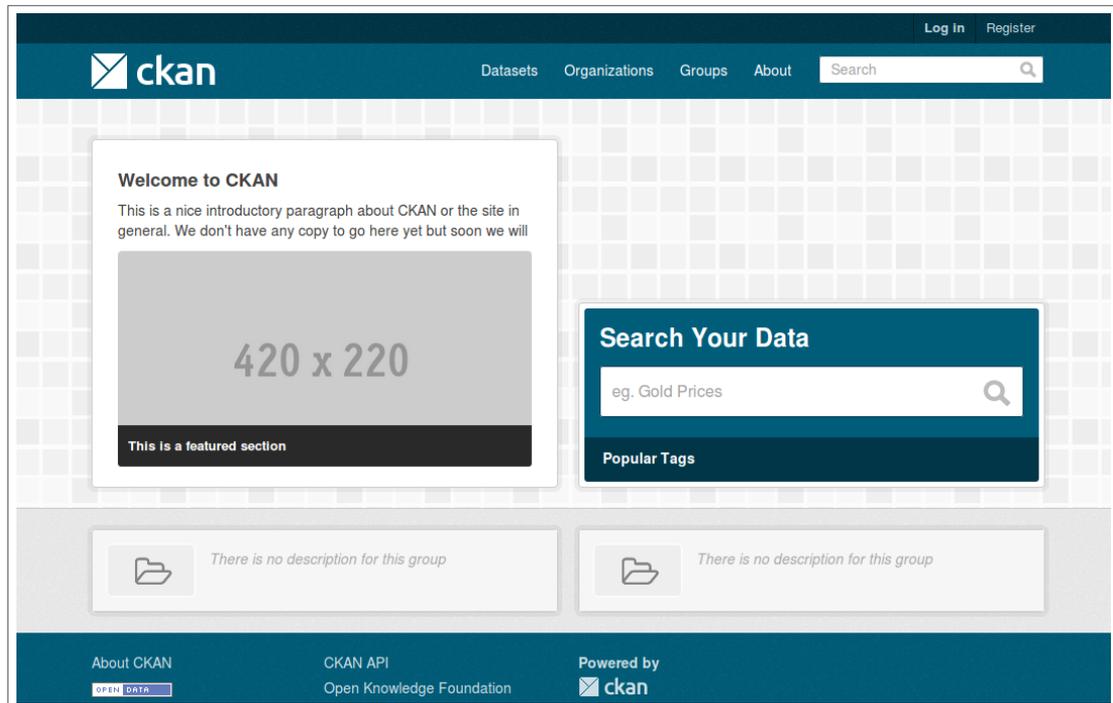
Serviço	Porta	Finalidade
NGINX	80	Proxy
uWSGI	8080	Servidor WEB
uWSGI	8800	DataPusher
Solr/Jetty	8983	Ferramenta de Busca
PostgreSQL	5432	SGBD
Redis	80	Servidor de Dicionário Remoto

Fonte: Adaptado de DOCS.CKAN (2022a)

site, e o conteúdo da seção “Sobre”. Os campos permitem sintaxe Markdown. O arquivo de configuração, por outro lado, serve para que sejam ajustados os seguintes parâmetros:

- URL e porta da aplicação.
- Nome de usuário e base de dados principal.
- Nome de usuário e base de dados do plugin DataStore.
- URL do Solr.
- URL do Datapusher.
- Caminho das imagens (logo e favicon).

Figura 8 – Página Inicial do portal



Fonte: Autoria Própria (2023)

- Idioma principal e opções adicionais.
- Tamanho máximo (em Megabytes (MB)) permitido para arquivos e imagens.
- Formatos permitidos para arquivos.
- Configurações do serviço de e-mail.

É nesta etapa que pode ser feita a restrição de formatos de arquivos que serão aceitos para armazenamento no DataStore. Seguindo os princípios de dados abertos, as opções inseridas foram CSV, JSON e Portable Document Format (PDF), sendo o último apenas para publicação dos dicionários de dados. Quanto ao tamanho, o valor padrão de 10 MB se mostrou insuficiente, sendo alterado para 20. O serviço de e-mail é utilizado unicamente para a validação e redefinição de senhas dos usuários. O protocolo Simple Mail Transfer Protocol (SMTP) deve ser configurado com a porta 587.

### 5.2.3 Configurações avançadas

Apesar de o CKAN oferecer várias opções de personalização, algumas alterações ficam restritas ao código do sistema. Nesta Seção, serão detalhadas as alterações - que incluem correções e melhorias de usabilidade feitas na aplicação.

Começando com os ajustes no código do CKAN, o primeiro deles ocorreu no texto das

mensagens de criação de usuário e redefinição de senha. O título e conteúdo do e-mail é definido no código, no idioma padrão, inglês. As mensagens foram traduzidas para o português.

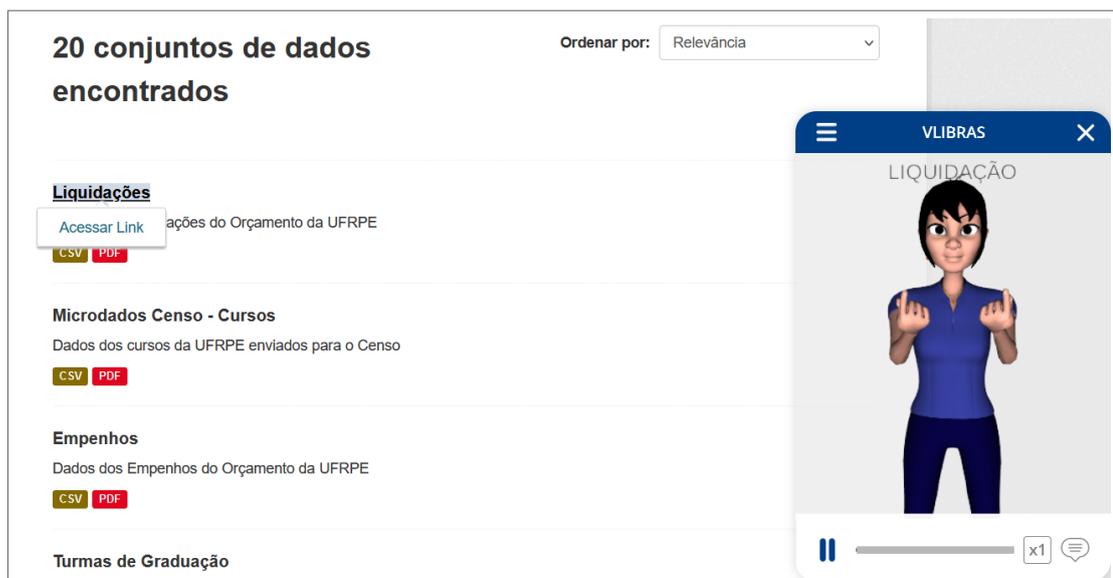
Em seguida, o cabeçalho e rodapé do site passaram por modificações para melhorar a usabilidade. No cabeçalho, os links “*Conjuntos de Dados*”, “*Organizações*” e “*Grupos*” foram reordenados, respeitando a seguinte hierarquia: primeiro Organizações, que possuem Grupos, que contém Conjuntos de Dados. Já no rodapé foram inseridos os links para o site oficial da UFRPE e para as páginas “*Sobre*” e “*Estatísticas*” do portal.

Na página principal, a imagem de fundo foi redimensionada para melhor se adequar ao conteúdo. A seção de destaques também foi alterada, recebendo uma imagem da instituição e legenda, além da inserção de um link que encaminha o usuário para o perfil da organização.

Por padrão a aplicação permite optar entre algumas opções de visualização, podendo exibir o campo de busca, a seção de estatísticas ou os conjuntos de dados em destaque. Um novo layout foi desenvolvido com base nessas alternativas, concentrando as principais funcionalidades do sistema em sua página inicial.

Para atender aos requisitos de acessibilidade, foi feita a instalação do *VLibras Widget*, que consiste em inserir a referência para a ferramenta dentro do código HTML do CKAN. A Figura 9 mostra o recurso em funcionamento dentro do portal. Já a Figura 10 exibe a página inicial do portal - disponível no endereço <http://dados.ufrpe.br/> - após as alterações aqui descritas.

Figura 9 – Utilização do VLibras no CKAN



Fonte: Autoria Própria (2023)

Figura 10 – Página Inicial do portal após personalização



Fonte: Autoria Própria (2023)

#### 5.2.4 Instalação e Configuração de *Plugins*

O CKAN possui diversos *plugins* que podem ser utilizados para personalizar a visualização dos *datasets*, sendo alguns já incluídos no pacote de instalação padrão e outros que devem ser inseridos à parte (DOCS.CKAN, 2022b). Esta Seção detalha quais destes *plugins* foram utilizados e o propósito de cada um deles.

Quadro 8 – Lista de *plugins* instalados

Nome	Incluso no CKAN	Finalidade
DataStore	Sim	Armazenamento de dados
DataPusher	Sim	Upload de dados para o DataStore
DataStore Grid	Sim	Visualização de tabelas
DataStore Graph	Sim	Visualização de Gráficos
Data Explorer	Sim	Visualização de tabelas
Text view	Sim	Visualização de arquivos de texto
Image view	Sim	Visualização de imagens
DataStore Map	Sim	Visualização de mapas interativos
PDF viewer	Não	Visualização de PDFs
Stats	Sim	Estatísticas

Fonte: Autoria Própria (2023)

Um dos *plugins* mais relevantes, que é pre-requisito para alguns dos serviços da aplicação, é o DataStore. A extensão armazena o conteúdo dos *datasets* em uma base de dados *ad hoc* -

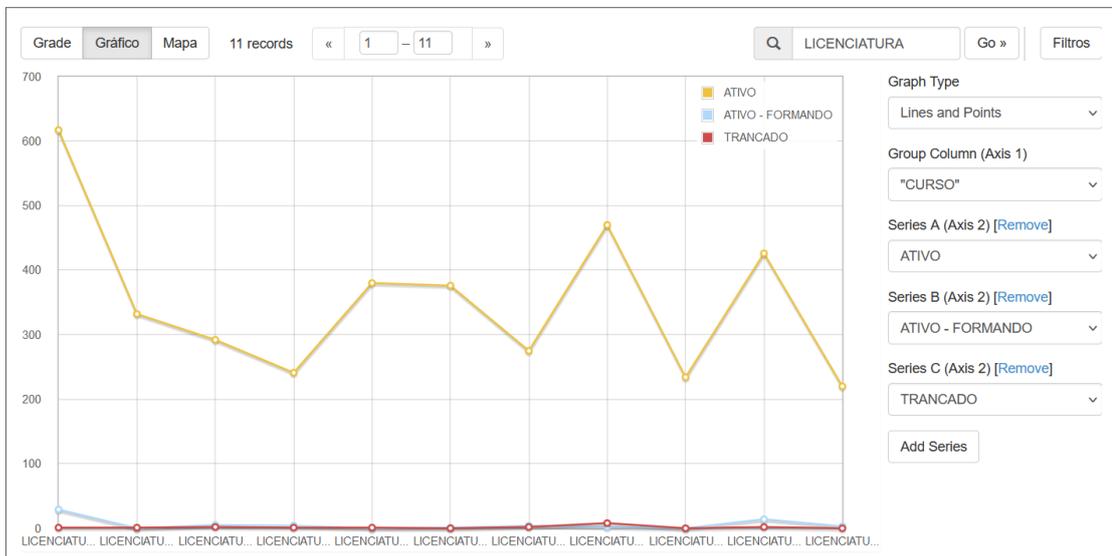
própria para este fim - e possibilita a pré-visualização dos recursos através do DataStore Grid, como tabelas, e o DataStore Graph, como gráficos. O DataStore também fornece uma API para leitura, busca e filtragem dos conjuntos de dados. Nas Figuras 11 e 12 é possível ver o mesmo conjunto de dados - quantitativo de alunos de graduação - com o filtro para os cursos de licenciatura e visualizações como tabela e gráfico, respectivamente.

Figura 11 – Plugin para visualização de dados em tabela

"CURSO"	ATIVO	ATIVO - FORMANDO	TRANCADO
LICENCIATURA EM CIÊNCIAS BIOLÓGICAS	617	29	1
LICENCIATURA EM COMPUTAÇÃO	332	0	1
LICENCIATURA EM EDUCAÇÃO FÍSICA	292	5	2
LICENCIATURA EM FÍSICA	241	4	1
LICENCIATURA EM HISTÓRIA	380	0	1
LICENCIATURA EM LETRAS - PORTUGUÊS E ESPANHOL	376	1	0
LICENCIATURA EM LETRAS - PORTUGUÊS E INGLÊS-UAST	275	4	2
LICENCIATURA EM MATEMÁTICA	470	2	8
LICENCIATURA EM PEDAGOGIA	234	0	0
LICENCIATURA EM QUÍMICA	426	14	2
LICENCIATURA EM QUÍMICA - UAST	220	3	0

Fonte: Autoria Própria (2023)

Figura 12 – Plugin para visualização de dados como gráfico



Fonte: Autoria Própria (2023)

O plugin funciona em conjunto com o DataPusher, que insere automaticamente os *datasets* na DataStore. Ambos já estão inclusos na instalação do CKAN por pacotes, sendo necessário apenas configurar uma nova base de dados no PostgreSQL.

Outra opção de plugin para visualização de dados é o Data Explorer, que não requer que o dataset esteja no DataStore. Isso é feito através de um recurso externo, o DataProxy, que

converte e retorna o conteúdo para visualização. Em comparação com o DataStore Grid, ele suporta menos formatos de arquivos e codificações.

Quando utilizados em conjunto, o Data Explorer serve como uma redundância para o DataStore Grid, oferecendo visualização de dados caso o serviço do DataStore/DataPusher esteja indisponível. Para isso, no arquivo de configuração principal da aplicação, existe uma seção em que são definidas as visualizações criadas por padrão.

Em adição à visualização em gráficos e tabelas para arquivos em CSV, o CKAN possui extensões para arquivos de texto, imagens e gráficos. São eles o Text view, que mostra o conteúdo de arquivos JSON ou XML com a sintaxe realçada; Image view, que exibe a imagem e insere uma tag com referência para sua Uniform Resource Locator (URL); e DataStore Map, que gera um mapa interativo através de valores de latitude e longitude.

O último dos *plugins* para exibição de conteúdo é o PDF Viewer, que renderiza os arquivos em PDF para pré-visualização. O recurso, que precisa ser inserido separadamente e se encontra disponível no GitHub, se mostrou uma boa alternativa para exibir os dicionários de dados dos *datasets*. A Figura 13 mostra um exemplo desta utilização.

Figura 13 – Plugin para visualização de documentos PDF



Fonte: Autoria Própria (2023)

Na parte de estatísticas de utilização, o CKAN oferece nativamente a extensão Stats. Voltada principalmente para portais com um grande número de usuários e *datasets*, ela fornece tabelas e gráficos com dados que incluem a quantidade de conjuntos de dados, frequência

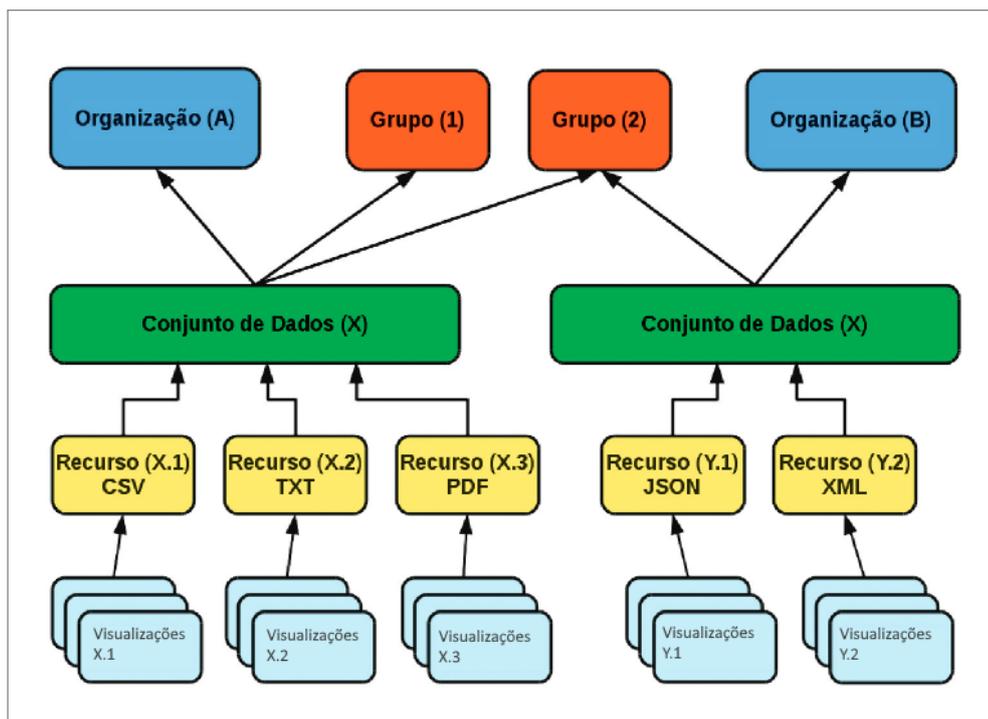
média de atualização, *datasets* mais editados e mais bem cotados, tags mais populares, maiores grupos e usuários que mais contribuem.

### 5.2.5 Integração com o Portal Brasileiro de Dados Abertos

O próximo passo no processo de implantação do portal se deu através da criação das Organizações e Grupos, além da personalização das páginas do CKAN de acordo com as necessidades da instituição. Na estrutura da aplicação, os conjuntos de dados pertencem a uma organização. Para catalogação no Portal Brasileiro de Dados Abertos, deve ser criada apenas uma organização por órgão ou entidade.

Os conjuntos de dados - que podem ser públicos ou privados - são compostos pelos seus metadados e recursos. Eles são agrupados por Organização e por Grupos temáticos, conforme detalhado na Figura 14. Cada conjunto pertence sempre a uma Organização e a um ou mais Grupos. Os metadados se aplicam para todos os recursos daquele conjunto e permitem o cadastro de descrição, autor, versão, licença, entre outros. Os recursos podem ter formatos diversos - como planilhas, imagens ou dados geográficos - e receber várias opções de visualização, permitindo a análise do conteúdo sem necessidade de baixar os arquivos.

Figura 14 – Agrupamento dos conjuntos de dados



Fonte: COSTA et al. (2017)

Já os grupos reúnem conjuntos de dados e devem refletir o que foi definido no PDA da instituição. Sendo assim, cada base prevista no documento - como Orçamento, Ensino de Graduação e Contratos - deu origem a um grupo no portal, como ilustrado na Figura 15. Para facilitar a identificação, foram inseridas imagens personalizadas para cada grupo.

Figura 15 – Organização de grupos no portal



Fonte: Autoria Própria (2023)

Conforme estabelecido na Seção 5.1, um dos fatores determinantes para a escolha do CKAN como aplicação para o portal de dados abertos da UFRPE foi a possibilidade de fazer a integração de forma automática com o portal do Governo Federal. O processo de sincronização do conteúdo - feita sempre que um recurso é inserido ou atualizado - é realizado pela CGU, que disponibiliza um canal de contato para as instituições que tenham interesse em realizar a integração. Qualquer órgão cujo portal de dados abertos utilize o CKAN pode solicitar a adesão (BRASIL, 2022a). Na Figura 16 é exibido o perfil da UFRPE no portal nacional, funcionando de forma automática.

Figura 16 – Página da UFRPE no portal GOV.BR



Fonte: Autoria Própria (2023)

### 5.3 Segurança

Durante o processo de implantação do portal, foi identificada uma vulnerabilidade do CKAN no formulário de cadastro. Nativamente, a aplicação não possui nenhum tipo de segurança extra para criação de usuários, como Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) ou e-mail de confirmação. Em um período de dois dias, foi registrada a criação de cerca de 700 usuários por meios automatizados, os chamados “bots”. Esse tipo de ataque é particularmente comum em *websites* que possuem cadastro gratuito (MORADI; KEYVANPOUR, 2015), como é o caso de portais de dados abertos.

Tal ocorrência é resultado de um ataque Distributed Denial of Service (DDoS) na camada de aplicação, que consiste no envio de uma grande quantidade de requisições para um serviço com o objetivo de torná-lo indisponível. Essa ameaça é particularmente difícil de ser enfrentada devido à dificuldade de identificar esse tipo de tráfego dos usuários legítimos (SHEA; LIU, 2012).

Duas soluções para esse problema foram exploradas. A primeira consistia na utilização do plugin CKAN TextCaptcha<sup>6</sup>, que insere uma validação por CAPTCHA no processo de criação de usuário. A segunda alternativa era remover a opção de criação de usuários, restringindo o registro para os usuários administradores do portal.

Apesar de ser uma solução eficiente para prevenir esse tipo de ataque, o uso de CAPTCHA possui algumas desvantagens. Sua principal inconveniência consiste em afetar negativamente a experiência dos usuários, piorando o fluxo de interação com a aplicação e muitas vezes

<sup>6</sup> Disponível em: <https://extensions.ckan.org/extension/textcaptcha/>. Acesso em: 18 abr. 2023.

fazendo com que o usuário desista de acessar a página. Outro ponto fraco dessa solução está na acessibilidade. A maior parte dos CAPTCHAs são inacessíveis ou extremamente difíceis para pessoas com deficiência visual ou visão comprometida (MORADI; KEYVANPOUR, 2015).

Por esse motivo e devido ao fato de o usuário registrado mas sem papel atribuído não possuir nenhum privilégio em relação ao usuário não logado (COSTA et al., 2017), a opção de restringir a criação de usuários pela interface WEB no arquivo de configuração principal do CKAN, mostrada no Apêndice A, foi considerada mais apropriada.

#### **5.4 Gerenciamento de permissões no CKAN**

O gerenciamento de permissões no portal de dados abertos está ligado a dois requisitos da legislação: fornecer acesso irrestrito às informações e garantir sua autenticidade e integridade. A configuração das organizações e grupos, descrita na Seção 5.2, Subseção 5.2.5, é essencial para essa etapa, uma vez que o papel do usuário e, conseqüentemente, as permissões a ele atribuídas são de acordo com a organização ou grupo ao qual ele pertence. A aplicação possui cinco tipos de usuário:

1. Usuário anônimo: usuário sem cadastro no sistema. Pode acessar todos os conjuntos de dados públicos.
2. Usuário registrado: usuário com cadastro, mas sem papel atribuído. Pode acessar todos os conjuntos de dados públicos. Removido conforme Seção 5.3.
3. Membro: Pode ver os conjuntos de dados privados da organização, mas não adicionar novos conjuntos de dados.
4. Editor: Pode adicionar e editar conjuntos de dados, mas não gerenciar membros da organização.
5. Administrador: Pode adicionar/editar e excluir conjuntos de dados, assim como gerenciar membros da organização.

O papel de administrador do sistema deve ser restrito aos gestores de dados abertos e pela Tecnologia de Informação (TI) da instituição, gerenciando toda a organização. No caso da UFRPE, seriam os membros do CTDA e da STD. Os responsáveis por cada base de dados devem ter papel de editor para o seu grupo. Já o público acessa o portal de forma anônima, tendo total liberdade de acesso aos dados disponibilizados.

Para garantir que as definições de organização e grupos, realizadas conforme o PDA, sejam respeitadas, foram feitos ajustes no arquivo de configuração principal do CKAN para

que apenas usuários com papel de administrador possam adicionar ou excluir essas entidades. A opção de criar conjuntos de dados não associados a nenhum grupo também foi removida.

## 5.5 Orientações para implementação do portal de dados abertos

Nesta Seção são resumidas as descobertas feitas ao decorrer do Capítulo, começando pela escolha da aplicação para gerenciamento de dados abertos. No caso de IFES, o CKAN não só atende a todos os requisitos da legislação, como também é o único que vai permitir a integração automática com o Portal Brasileiro de Dados Abertos.

No entanto, as instituições devem considerar que a utilização do CKAN não garante a adesão total à legislação. Plugins para visualização e disponibilização de API são um requisito essencial, assim como a utilização do VLibras para cumprir os requisitos de acessibilidade. O mesmo pode ser dito dos ajustes de segurança, para evitar ataques DDoS.

Já as melhorias no *layout* para usabilidade, apesar de não serem essenciais, contribuem para a experiência do usuário, e conseqüentemente para o sucesso da iniciativa de dados abertos da instituição. Os ajustes nas permissões, por sua vez, garantem que a estrutura das Organizações e Grupos será respeitada, e que os conjuntos de dados serão criados corretamente em seus respectivos Grupos, por usuários autorizados.

Outro ponto importante consiste em disponibilizar uma referência ao portal de dados abertos no site institucional. Para garantir o cumprimento da LAI, uma seção específica, denominada “*Acesso à Informação*”, deve constar no menu principal das páginas oficiais dos órgãos. Esta página de “*Acesso à Informação*” deve obedecer a estrutura definida no GTA e conter o link para o portal de dados abertos. A adesão ao formato e nomenclatura propostos facilita a localização da informação pelo cidadão (BRASIL, 2019b).

## 6 ESTUDO DE CASO

O presente Capítulo apresenta um estudo de caso, com o objetivo mapear as etapas do processo de abertura de dados pelos responsáveis. Para isso, foram escolhidas três bases de dados do PDA da UFRPE: Ensino de Graduação, Financeiro/Orçamento e Contratos. A seleção, que também levou em consideração a relevância das bases para as IFES e ordem de prioridade no PDA, buscou englobar dados em diferentes estágios.

O critério para seleção dos participantes do estudo de caso foi o próprio PDA que, por determinação da INDA, contém a lista das bases e o responsável por sua abertura. No processo de aprovação do documento, esses responsáveis são notificados, tendo ciência de sua participação no processo. Os encontros e reuniões foram organizados pelo CTDA, no papel de responsável pela gestão dos dados abertos da instituição.

O processo de publicação dessas bases teve início através de um diálogo com os responsáveis pelos dados, a fim de avaliar a melhor forma de extrair as informações dos sistemas informatizados. Em seguida, esses dados foram adaptados para o formato aberto e foi produzido um modelo de dicionário de dados. Esses artefatos foram aprovados pelos responsáveis pelas bases e disponibilizados no portal.

O presente Capítulo está organizado da seguinte forma: a Seção 6.1 relata o processo de abertura dos dados de Ensino de Graduação e justifica sua divisão em dois grupos; a Seção 6.2 demonstra como foi feita a publicação dos dados da execução orçamentária e financeira; a Seção 6.3 traz este processo com os contratos da instituição; e a Seção 6.4 sintetiza as experiências do estudo de caso.

### 6.1 Caso I - Ensino de Graduação

A base de Ensino de Graduação está em terceiro lugar na ordem de prioridade no PDA da UFRPE e foi escolhida para ser o foco deste primeiro estudo de caso. No documento, os órgãos responsáveis pelas informações são identificados como sendo a Pró-Reitoria de Planejamento (PROPLAN) e a Pró-Reitoria de Ensino de Graduação (PREG). A PROPLAN fica a cargo das informações enviadas anualmente para o Censo da Educação Superior do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), enquanto a PREG gere as demais informações relacionadas aos cursos de graduação da instituição.

Por esse motivo, foram criados dois grupos no CKAN: “*Censo da Educação Superior*” e “*Ensino de Graduação*”. Essa separação permite que cada responsável seja o único editor de seus dados e tenha acesso irrestrito para adicionar, remover e editar conjuntos de dados.

Ao contrário de outras bases, em que os dados a serem divulgados ficaram a critério do servidor identificado no PDA, para graduação o documento já prevê a abertura de alguns dados específicos. Entre eles, pode-se destacar dados dos cursos de graduação a distância; quantitativo de alunos de graduação; lista das matrizes curriculares da graduação; turmas de graduação; disciplinas de graduação; matriculados nas turmas de graduação; censo da educação superior; lista dos docentes da instituição.

Além dos dados previstos, os editores optaram por acrescentar informações de interesse da comunidade. No caso dos dados geridos pela PREG, foram acrescentados os dados dos coordenadores de cursos e eventuais substitutos e a lista de componentes por perfil curricular.

#### 6.1.1 Dados do sistema acadêmico

A primeira etapa do estudo de caso teve como objetivo realizar a abertura dos dados de Ensino de Graduação, explorando o processo de extração, tratamento e publicação das informações no portal de dados abertos da instituição.

O sistema acadêmico utilizado pela UFRPE é o Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), que faz parte dos Sistemas Institucionais Integrados de Gestão (SIGs), desenvolvidos pela Universidade Federal do Rio Grande do Norte (UFRN). Como a aplicação não possibilita a extração de relatórios com a informação desejada em formato de planilha, as informações a serem publicadas no portal de dados abertos tiveram que ser extraídas diretamente da base de dados. Conforme estabelecido na Seção 4.2 do Capítulo 4, o SGBD utilizado pelo sistema é o PostgreSQL. O acesso ao banco foi feito através do *software* DBeaver. As consultas, feitas em SQL, foram exportadas em formato CSV.

Para simplificar as consultas e evitar repetição de código, foi criada uma função auxiliar que realiza o filtro dos cursos de graduação ativos na instituição. Tal função recebe o código do curso e verifica sua situação. Caso o curso se enquadre em algum critério de inatividade, como não possuir nenhum aluno, não possuir código INEP ou estar em uma estrutura organizacional extinta, é retornado um valor booleano indicando a situação, a ser utilizado nas demais consultas. Esse passo se mostrou necessário pois nem todos os cursos inativos estão marcados como tal no sistema. Um trecho dessa função pode ser visualizado na Figura 17.

Também foram utilizadas funções nativas do PostgreSQL, tais como: *CROSSTAB* - rotaciona o resultado da consulta para exibição de linhas como colunas; *UNNEST* - expande um array em um conjunto de linhas; *ARRAY\_AGG* - recebe um conjunto de valores e retorna um *array*; além de outras funções frequentemente utilizadas em consultas SQL, tais como *TO\_CHAR*, *NOW*, *MAX* e *COUNT*.

Figura 17 – Criação de função auxiliar para filtro de cursos ativos

```
1 CREATE OR REPLACE FUNCTION public.cursos_ativos(id int)
2 RETURNS boolean
3 LANGUAGE plpgsql
4 AS $function$
5 DECLARE
6     ativo boolean;
7 BEGIN
8     IF (--filtro para cursos inativos)
9         THEN ativo = false;
10        ELSE ativo = true;
11    END IF;
12    RETURN ativo;
13 END;
14 $function$
15 ;
16
```

Fonte: A autoria Própria (2023)

Por orientação da INDA, os dados devem estar em seu estado primário, ou seja, com o mínimo de agregações possíveis. Isso permite que os usuários possam realizar suas próprias análises, não ficando restritos ao que foi disponibilizado pela instituição. Agregações podem ser divulgadas, desde que de forma adicional.

Esta foi a lógica adotada para a extração dos dados básicos como de cursos de graduação, matrizes curriculares, componentes e docentes. As únicas combinações realizadas foram feitas com o objetivo de extrair descrições como nome do município ou curso, de forma a não exibir apenas o código. A nomenclatura das colunas foi ajustada através do uso de *aliases* para facilitar o entendimento por parte do público. A Figura 18 exibe um exemplo de consulta para retornar os cursos da instituição, com uso da função auxiliar criada para filtrar os ativos.

No caso das bases de turmas de graduação, foi necessário fazer a divisão pela dimensão temporal. Devido à natureza semestral da oferta, optou-se por publicar as bases por semestre. Para publicação no portal de dados abertos, o CKAN permite a atualização dos arquivos, no caso dos que dispensam esta divisão por dimensão temporal, ou que sejam adicionados diversos arquivos por conjunto de dados, um por dimensão, como pode ser visto na Figura 19.

Figura 18 – Pseudocódigo de consulta para extração de dados de graduação

```

1 select c.nome as "NOME_CURSO",
2 m.nome as "MUNICIPIO",
3 c.codigo_inep as "CODIGO_INEP",
4 from curso c
5 inner join municipio m
6 [...]
7 where public.cursos_ativos(c.id_curso) is true
8
9
10

```

Fonte: Autoria Própria (2023)

Figura 19 – Conjunto de dados de Turmas de Graduação no CKAN

**Turmas de Graduação**

Dados das turmas de graduação ofertadas por semestre.

**Dados e recursos**

- Dicionário de Dados - Turmas de graduação** (PDF)  
Descrição das colunas da base turmas de graduação. [Explorar](#)
- Turmas de Graduação - 2020.1** (CSV)  
Dados das turmas de graduação ofertadas no semestre. [Explorar](#)
- Turmas de Graduação - 2020.2** (CSV)  
Dados das turmas de graduação ofertadas no semestre. [Explorar](#)
- Turmas de Graduação - 2021.1** (CSV)  
Dados das turmas de graduação ofertadas no semestre. [Explorar](#)
- Turmas de Graduação - 2021.2** (CSV)  
Dados das turmas de graduação ofertadas no semestre. [Explorar](#)
- Turmas de Graduação - 2022.1** (CSV)  
Dados das turmas de graduação ofertadas no semestre. [Explorar](#)

Fonte: Autoria Própria (2023)

Durante a elaboração das listagens de alunos dos cursos de graduação e matriculados nas turmas de graduação, optou-se por utilizar quantitativos. Esta decisão foi tomada como forma de respeitar à LGPD, uma vez que sem o tratamento estas listagens conteriam dados pessoais dos discentes. Não devem ser exibidos nem dados pessoais nem identificadores que, em um eventual cruzamento de dados, permitam a identificação destas informações pessoais.

A Figura 20 mostra a consulta feita para extração do quantitativo de alunos por curso. Nela, a função *COUNT()* foi utilizada para exibir o total de alunos por situação. Em seguida, a função *crosstab* manipula o resultado para que a exibição do total de alunos por situação seja em colunas, ao invés de linhas. Assim como com as turmas, os dados foram divididos pela

dimensão temporal, por semestre. O resultado é mostrado na Figura 21.

Figura 20 – Pseudocódigo de consulta para quantitativo de alunos

```

1 SELECT *
2 FROM crosstab(
3   $$select c.nome,
4   d.descricao,
5   count(d.descricao) as QUANTIDADE
6   from discente d
7   inner join curso c
8   [...]
9   group by c.nome, d.descricao
10  order by 1$$
11  , $$SELECT unnest('{ATIVO,"ATIVO - FORMANDO",TRANCADO}'::text[])$
12  ) AS ct ("CURSO" text, "ATIVO" int, "ATIVO - FORMANDO" int, "TRANCADO" int);
13
14

```

Fonte: Autoria Própria (2023)

Figura 21 – Quantitativo de alunos de graduação no CKAN



Quantitativos de alunos de graduação - 2022.1

URL: <http://dados.ufpe.br/dataset/e85ec350-5a10-440f-8c7c-4e4888cbd580/resource/d386247f-4a97-428e-adb3-2cf2baa88ffc/download/alunos.csv>

Quantitativos de alunos de graduação, por curso e situação no semestre 2022.1.

Explorador de Dados

Tela cheia Embutir

Grade Gráfico Mapa 45 records « 1 - 45 » Search data ... Go » Filtros

"CURSO"	ATIVO	ATIVO - FORMANDO	TRANCADO
BACHARELADO EM ADMINISTRAÇÃO	340	9	6
BACHARELADO EM ADMINISTRAÇÃO - UAST	221	3	0
BACHARELADO EM AGROECOLOGIA	82	0	1
BACHARELADO EM AGRONOMIA	424	4	4
BACHARELADO EM AGRONOMIA - UAST	243	2	1
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO	400	4	11
BACHARELADO EM CIÊNCIAS BIOLÓGICAS	325	1	3
BACHARELADO EM CIÊNCIAS BIOLÓGICAS - UAST	198	9	0
BACHARELADO EM CIÊNCIAS DO CONSUMO	249	3	1
BACHARELADO EM CIÊNCIAS ECONÔMICAS	304	2	2

Fonte: Autoria Própria (2023)

Para o quantitativo matriculados nas turmas de graduação, a função *count()* foi usada em conjunto com a expressão *case* para extrair o total de matriculados por situação acadêmica. O código agrupa situações que tem o mesmo efeito para o estudante, mas internamente recebem códigos diferentes. É o caso de aprovação, que pode estar como aprovação por nota e aprovação por média, e dispensa, que pode ser do tipo dispensa ou aproveitamento. A solução utilizada facilita a leitura e entendimento por parte do público. Um pseudocódigo pode ser visto na Figura 22. Já a Figura 23 exhibe o resultado em um gráfico gerado pelo CKAN, com filtro para a disciplina de código 02106. É possível analisar por exemplo a diferença entre as situações de aprovação e reprovação de um mesmo componente, de acordo com a turma.

Figura 22 – Pseudocódigo de consulta para quantitativo de matrículas

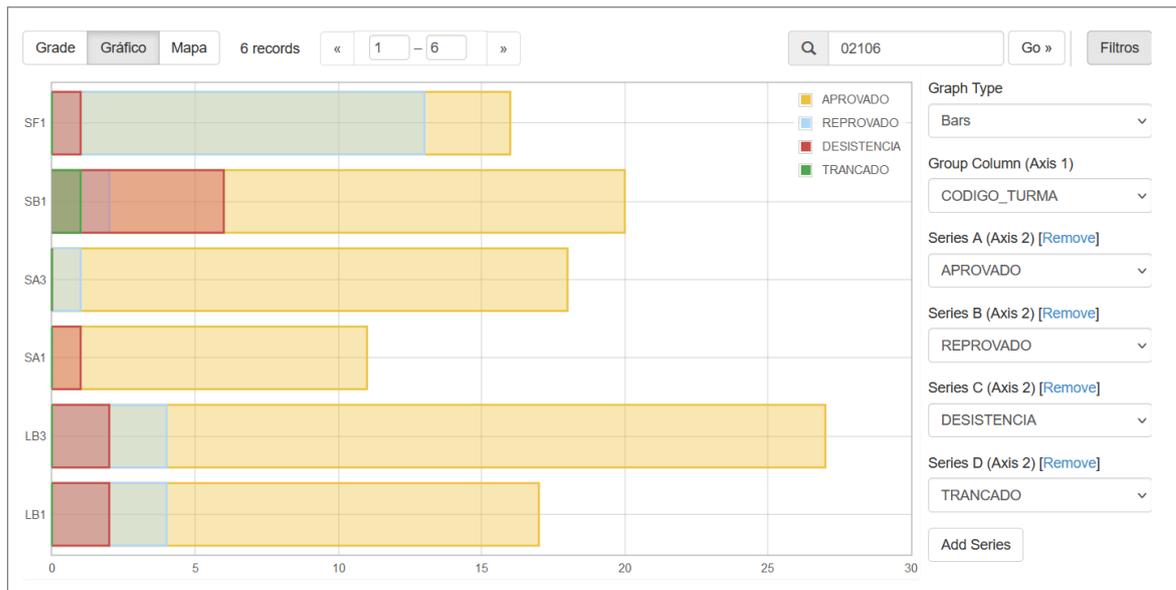
```

1 count(case when id_situacao in (aprovado) then matricula end) as "APROVADO",
2 count(case when id_situacao in (reprovado) then matricula end) as "REPROVADO",
3 count(case when id_situacao in (dispensado) then matricula end) as "DISPENSADO",
4 count(case when id_situacao in (trancado) then matricula end) as "TRANCADO",
5

```

Fonte: Autoria Própria (2023)

Figura 23 – Quantitativos de matriculados nas turmas de graduação no CKAN



Fonte: Autoria Própria (2023)

Enquanto alguns dados pessoais devem ser removidos ou anonimizados, existem situações em que as instituições têm obrigação legal de disponibilizar informações que contém esse tipo de dado. No caso das IFES, essas exceções podem incluir as listas de docentes e coordenadores de curso. Nestas situações, o dicionário de dados deve incluir uma menção à hipótese legal que permite a divulgação dos dados pessoais sem consentimento do titular.

Além da anonimização dos dados pessoais, quando necessário, e da manipulação do resultado da consulta para exibir as informações como colunas, ao invés de linhas, facilitando a leitura das planilhas, o processo de transformação e limpeza dos dados incluiu formatação de datas para o padrão *DD/MM/AAAA* e utilização de *alias* nas colunas.

Outra questão que precisou ser tratada foi a periodicidade das informações. Como a graduação funciona em ciclos semestrais, as situações dos alunos e turmas muda ao longo do semestre. Idealmente os relatórios devem ser gerados entre um semestre e outro mas, para evitar inconsistências caso a extração seja feita no meio do semestre, foi preciso inserir restrições nas consultas para evitar exibir situações temporárias. A listagem do quantitativo de matriculados por turma de graduação, por exemplo, retorna apenas as turmas cuja situação

esteja consolidada no fim do último semestre letivo.

O processo de extração das informações do sistema acadêmico foi devidamente documentado, de forma a permitir extrações periódicas, conforme o prazo estabelecido no PDA da instituição. Os dados foram validados pelos responsáveis, que analisaram tanto a adequação à LGPD como a completude e relevância. Por fim, foi elaborado um dicionário de dados para cada base a ser publicado, com as descrições das colunas e tipo de dado.

O maior desafio encontrado na extração dos dados de Ensino de Graduação foi a quantidade de informações previstas para divulgação no PDA e o fato de o sistema acadêmico não permitir a geração de relatórios no modelo desejado. A coleta das informações exige um alto nível de conhecimento da estrutura do banco de dados, necessário tanto para a extração como para o tratamento. Assim, as instituições devem mensurar o esforço necessário para este procedimento e incluí-lo nos prazos do PDA, além de formalizar o envolvimento por parte da TI da instituição, tanto na extração e limpeza inicial dos dados quanto nas atualizações periódicas.

### 6.1.2 Dados do censo de educação superior

Esta etapa do estudo de caso abordou a inclusão dos dados do Censo da Educação Superior no portal de dados abertos. A proposta envolveu analisar juntamente com a área negocial a melhor forma de inserir os dados, que já eram tratados pela instituição, na plataforma, com os devidos cuidados com dados pessoais e sensíveis e sem gerar um esforço desnecessário de retrabalho para a equipe do setor.

O Censo da Educação Superior, realizado anualmente pelo INEP, é o principal instrumento de pesquisa do MEC em relação às instituições de ensino superior públicas e privadas. São recolhidas informações sobre os cursos e locais de oferta, docentes e discentes (INEP, 2022).

A pesquisa tem como objetivo fornecer estatísticas confiáveis sobre a educação superior no país, provendo meios de realizar atividades de acompanhamento e avaliação, programas de expansão e de melhoria da qualidade do ensino, assim como contribuir para o trabalho dos gestores das instituições e do governo (INEP, 2022).

Na UFRPE, o envio dos dados, que é obrigatório para todas as instituições (Decreto nº 6425, de 2008), é feito pela PROPLAN. Do relatório enviado para o Censo, foram extraídas as seguintes listagens: laboratórios, laboratórios por curso, bibliotecas, bibliotecas por local, locais de oferta, cursos de graduação e cursos por local.

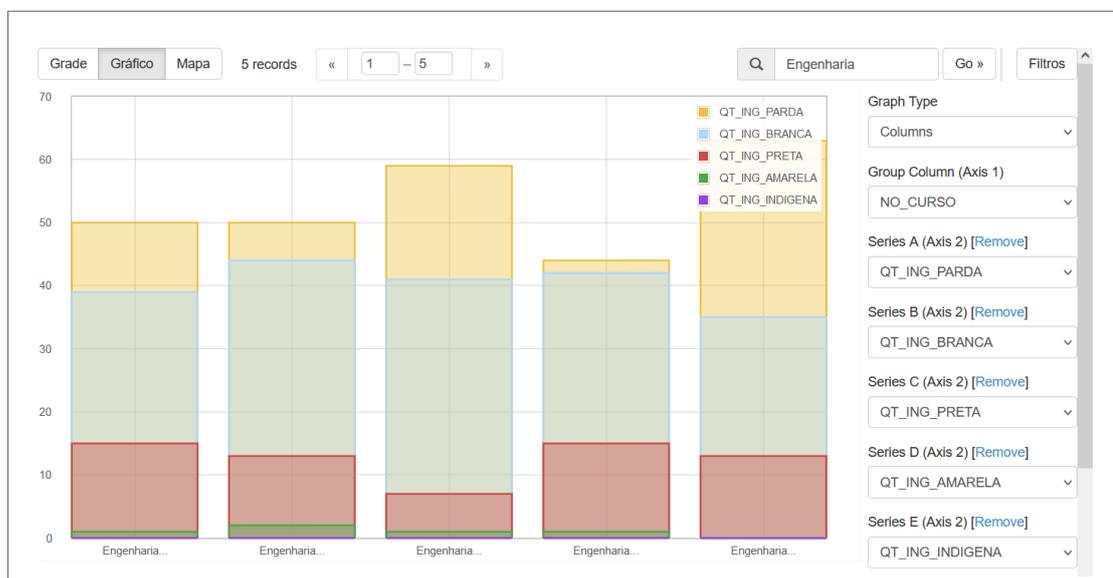
Para publicação destas listagens no portal de dados abertos, não foi necessário fazer ne-

nhum tipo de tratamento ou limpeza, uma vez que esse procedimento já é realizado para envio para o INEP. Além disso, os dados já se encontravam no formato CSV. O dicionário de dados foi produzido conforme as definições dos campos usada pelo próprio INEP.

Além das listagens mencionadas acima, a instituição também envia para o Censo as relações de alunos, docentes, docentes por curso e ingressantes. Por padrão, o Censo coleta informações como nome - dado considerado pessoal - raça e tipo de deficiência - que pela LGPD são classificados como sensíveis. Para o propósito de transparência ativa, a divulgação desses dados não se enquadra em cumprimento de obrigação legal.

Neste caso, o próprio INEP faz o tratamento dos dados antes de divulgá-los, usando a estratégia de agregação, mostrando apenas os quantitativos, de forma a impossibilitar a identificação dos indivíduos. Desta forma, a melhor alternativa para o portal de dados abertos foi utilizar as informações já formatadas pelo INEP, evitando retrabalho e garantindo o respeito à LGPD. A Figura 24 exibe umas das possíveis visualizações destes dados dentro do portal, com os quantitativos de ingressantes por raça nos cursos de Engenharia.

Figura 24 – Dados do Censo em visualização no CKAN



Fonte: Autoria Própria (2023)

Esta etapa do estudo de caso mostrou como artefatos já produzidos pela instituição podem acrescentar valor à iniciativa de dados abertos com o mínimo de esforço dos responsáveis. Também foi possível observar a necessidade de fazer o tratamento dos dados pessoais.

## 6.2 Caso II - Orçamento

Esta etapa do estudo de caso se propôs a investigar como se dá a publicação de dados vindos de sistemas do Governo Federal em casos em que as IFES já fazem um tratamento dessas informações. O objetivo era utilizar essas informações que já se encontravam devidamente tratadas para publicação no portal de dados abertos.

A gestão orçamentária da UFRPE é feita pela Gerência de Contabilidade e Finanças (GCF), ligada à Pró-Reitoria de Administração (PROAD). Além de já divulgar os dados da execução orçamentária e financeira no Portal da Transparência, essas informações já eram disponibilizadas no próprio site do setor, em um painel de monitoramento mostrado na Figura 25.

Figura 25 – Painel de execução orçamentária da GCF

Composição Atual do Orçamento UFRPE 2022							
Res. Lei	Dot. Inicial	Dot. Atualizada	Resultado Orçamentário	Créd. Disponível	Créd. Bloqueado		
<b>PRIMARIO OBRIGATORIO</b>	<b>R\$ 503.912.080,00</b>	<b>R\$ 511.737.482,00</b>	<b>R\$ 7.825.402,00</b>	<b>R\$ 6.363.895,27</b>	<b>R\$ 0,00</b>	<b>R\$ 649.073.904,00</b>	
20TP	R\$ 313.348.638,00	R\$ 315.013.776,00	R\$ 1.665.138,00	R\$ 1.957.720,41	R\$ 0,00	Dot. Inicial	
0181	R\$ 164.454.615,00	R\$ 171.875.299,00	R\$ 7.420.684,00	R\$ 1.389.943,39	R\$ 0,00	R\$ 656.031.880,00	
2128	R\$ 15.599.933,00	R\$ 15.599.933,00	R\$ 0,00	R\$ 2.213.843,56	R\$ 0,00	Dot. Atualizada	
2004	R\$ 4.666.191,00	R\$ 4.666.191,00	R\$ 0,00	R\$ 801.385,38	R\$ 0,00	R\$ 6.957.976,00	
0005	R\$ 5.841.703,00	R\$ 4.581.283,00	-R\$ 1.260.420,00	R\$ 2,53	R\$ 0,00	Resultado Orçamentário	
0056	R\$ 1.000,00	R\$ 1.000,00	R\$ 0,00	R\$ 1.000,00	R\$ 0,00	R\$ 11.148.038,14	
<b>FINANCEIRO</b>	<b>R\$ 83.689.109,00</b>	<b>R\$ 83.689.109,00</b>	<b>R\$ 0,00</b>	<b>R\$ 3.900.182,19</b>	<b>R\$ 0,00</b>	<b>R\$ 0,00</b>	
09HB	R\$ 83.689.109,00	R\$ 83.689.109,00	R\$ 0,00	R\$ 3.900.182,19	R\$ 0,00	Créd. Disponível	
<b>Total</b>	<b>R\$ 649.073.904,00</b>	<b>R\$ 656.031.880,00</b>	<b>R\$ 6.957.976,00</b>	<b>R\$ 11.148.038,14</b>	<b>R\$ 0,00</b>	<b>R\$ 0,00</b>	
<b>Composição Atual do Orçamento UFRPE 2022</b>							
PTRES	ID Fonte Recursos	ID Nat. Despesa	Dot. Inicial	Dot. Atualizada	Suplementação Orçamentária	Créd. Disponível	Créd. Bloqueado
087333	8100000000	319000	R\$ 1.273.232,74	R\$ 1,74	-R\$ 1.273.231,00	R\$ 1,74	R\$ 0,00
087333	8100000000	319091	R\$ 4.410.754,26	R\$ 4.410.754,26	R\$ 0,00	R\$ 0,00	R\$ 0,00
087333	8100000000	339000	-R\$ 12.810,21	R\$ 0,79	R\$ 12.811,00	R\$ 0,79	R\$ 0,00
087333	8100000000	339091	R\$ 170.526,21	R\$ 170.526,21	R\$ 0,00	R\$ 0,00	R\$ 0,00
137087	0150000000	335000	-R\$ 2.396,00	R\$ 0,00	R\$ 2.396,00	R\$ 0,00	R\$ 0,00
137087	0150262480	335041	R\$ 2.396,00	R\$ 2.396,00	R\$ 0,00	R\$ 0,00	R\$ 0,00
137088	0150000000	335000	-R\$ 14.427,37	R\$ 0,63	R\$ 14.428,00	R\$ 0,63	R\$ 0,00
137088	0150262480	335041	R\$ 14.427,37	R\$ 14.427,37	R\$ 0,00	R\$ 0,00	R\$ 0,00
<b>Total</b>			<b>R\$ 649.073.904,00</b>	<b>R\$ 656.031.880,00</b>	<b>R\$ 6.957.976,00</b>	<b>R\$ 11.148.038,14</b>	<b>R\$ 0,00</b>

Fonte: GCF (2022)

Nesta etapa do trabalho, o principal obstáculo encontrado estava na definição do PDA da UFRPE em separar a execução orçamentária e financeira em duas bases distintas. Juntamente com o servidor responsável pelos dados, foi estabelecido que essa divisão não faz sentido na prática. Por isso, foi criado apenas um grupo no CKAN denominado Orçamento e Financeiro, no qual foram publicadas as bases de Liquidações e Empenhos.

De forma a reduzir o retrabalho e incentivar que a atualização do portal de dados abertos fosse mantida de acordo com a periodicidade proposta no PDA - mensal -, os mesmos dados contidos no painel foram publicados no portal. Isso foi possível devido ao fato de a GCF já ter

realizado todo o trabalho de limpeza e tratamento dos dados previamente, sendo necessário apenas converter o formato do arquivo para CSV.

A produção do dicionário de dados também se mostrou complexa. Apesar de os dados já estarem prontos para publicação, a nomenclatura das colunas não permitia o entendimento e análise por usuários leigos. Assim, foi preciso inserir essas definições no dicionário, juntamente com o tipo de dado e obrigatoriedade, para possibilitar seu uso pela população.

Este estudo evidenciou a relevância das etapas de tratamento e limpeza no processo de abertura de dados. Como neste caso o repensável já realizava esse trabalho de forma rotineira, a inclusão no portal foi relativamente simples, não sendo necessário o envolvimento da TI da instituição. A Figura 26 exibe os conjuntos do grupo Orçamento e Financeiro.

Figura 26 – Grupo Orçamento e Financeiro no Portal

The screenshot displays the 'Orçamento e Financeiro' group page. At the top, there is a breadcrumb trail: 'Grupos / Orçamento e Financeiro'. Below this, there is a large icon representing a document with a dollar sign and a calculator. The main content area features a search bar with the placeholder text 'Pesquisar conjuntos de dados...'. Below the search bar, it states '2 conjuntos de dados encontrados'. The first result is 'Liquidações', described as 'Dados das Liquidações do Orçamento da UFRPE', with buttons for 'CSV' and 'PDF'. The second result is 'Empenhos', described as 'Dados dos Empenhos do Orçamento da UFRPE', also with 'CSV' and 'PDF' buttons. On the left side, there is a sidebar with the title 'Orçamento e Financeiro' and a description: 'Dados sobre a Execução orçamentária e financeira', followed by a 'Leia mais' link.

Fonte: Autoria Própria (2023)

### 6.3 Caso III - Contratos

A última etapa deste estudo se deu com a base de Contratos. A proposta era mapear os passos envolvidos na publicação de dados oriundos dos sistemas do Governo Federal mas que,

ao contrário do caso anterior, não passavam por nenhum tratamento dentro da instituição. Na UFRPE, a gestão de contratos é feita pela Coordenadoria de Acompanhamento Técnico Financeiro (CATF), da PROAD, através do sistema Contratos, módulo do *Compras.gov.br*, a solução do governo federal para controle orçamentário e contratual.

Como o PDA não especificava em detalhes os dados que deveriam ser publicados, a primeira etapa do processo consistiu na escolha das informações que seriam divulgadas no portal de dados abertos. A decisão, tomada juntamente com a CATF, foi pela relação dos contratos vigentes da instituição, em formato similar ao divulgado na seção de Acesso à Informação do site da UFRPE. Como o volume de dados não é tão grande, eles foram divididos na dimensão anual. Conforme definido no PDA, a atualização será mensal. A cada atualização, os recursos são sobrescritos pela versão mais nova.

Quanto à extração dos dados, a princípio o sistema Contratos já permite que ela seja feita em arquivo CSV. No entanto, foi necessário um trabalho extenso de limpeza e correção de inconsistências. Ao avaliar as informações extraídas do sistema, o responsável identificou vários campos que desnecessários ou que continham informações incorretas. Era o caso do número de parcelas e os dados da autoridade signatária, que sempre estão em branco.

Sendo assim, a CATF propôs uma adaptação do arquivo extraído do sistema, visando melhorar a qualidade da informação e facilitar o entendimento por parte do público. Foram inseridas informações sobre a unidade de origem da licitação e situação (ativo ou encerrado). Este modelo será utilizado para elaboração das próximas atualizações do arquivo.

Além do detalhamento das colunas, o dicionário de dados recebeu os possíveis valores de alguns dos campos utilizados, como categoria, natureza e termo do contrato. Devido à possibilidade de conter o nome e CPF dos fornecedores, o documento recebeu uma justificativa para que o mesmo ficasse de acordo com a LGPD, uma vez que nesse caso essa divulgação se enquadra na hipótese de cumprimento de obrigação legal. A Figura 27 mostra os detalhes do conjunto de dados de Contratos, com dois recursos, o arquivo CSV e o dicionário, além dos metadados padrão do CKAN: autor, mantenedor, datas e categoria do VCGE.

Essa etapa do estudo difere das anteriores principalmente no tratamento e limpeza dos dados. No caso dos dados de graduação, esse processo foi realizado durante a extração do banco de dados ou seguindo as especificações do Censo. Já os responsáveis pelo Orçamento já tinham as informações devidamente tratadas. Para Contratos, foi necessário estabelecer os critérios para essa limpeza, decidindo quais dados manter, corrigir ou eliminar, tendo sempre como critério a melhor experiência do usuário do portal.

Figura 27 – Base de Contratos com metadados no CKAN

**Dados e recursos**

**Dicionário de Dados - Contratos**  
Descrição dos campos da base de contratos. [Explorar](#)

**Contratos - 2022**  
Informações dos contratos vigentes em 2022. [Explorar](#)

compras contratos

**Informações Adicionais**

Campo	Valor
Autor	<a href="#">Dados Abertos UFRPE</a>
Mantenedor	<a href="#">Dados Abertos UFRPE</a>
Estado	active
Última Atualização	24 de novembro de 2022, 14:52 (UTC-03:00)
Criado	7 de novembro de 2022, 15:08 (UTC-03:00)
Categoria VCGE	Compras governamentais

Fonte: Autoria Própria (2023)

## 6.4 Principais descobertas

Esta Seção tem como objetivo reunir as principais descobertas do estudo de caso com a implementação do portal de dados abertos da UFRPE. O estudo começou com a disponibilização da aplicação em ambiente de produção, após o processo de configuração e personalizações do CKAN detalhado no Capítulo 5.

Neste momento, o PDA da instituição ainda estava em fase de elaboração, faltando os ajustes finais e aprovação pelo CONSU. Foi necessário iniciar o processo de abertura dos dados antes da publicação do documento para poder cumprir os prazos propostos pelo CTDA.

Foi possível observar a relevância da participação dos responsáveis pela publicação dos dados não só no momento de inserir as informações no portal, mas também na elaboração do PDA. Tanto em Graduação quanto em Orçamento foi preciso fazer ajustes no que o documento previa - a frequência de atualização e a denominação das bases de dados, respectivamente.

A pesquisa também mostrou os desafios das etapas de extração, tratamento e limpeza dos dados. O nível de complexidade pode variar bastante de acordo com a qualidade dos dados originais, do nível de conhecimento do responsável pela informação e da necessidade ou não de extrair as informações diretamente do banco de dados.

O estudo de caso mostrou as etapas necessárias para a abertura dos dados por parte das

IFES. Os problemas encontrados e suas respectivas soluções foram utilizados como base para elaboração de um guia de boas práticas para publicação de dados abertos.

## 7 BOAS PRÁTICAS PARA PUBLICAÇÃO DE DADOS ABERTOS

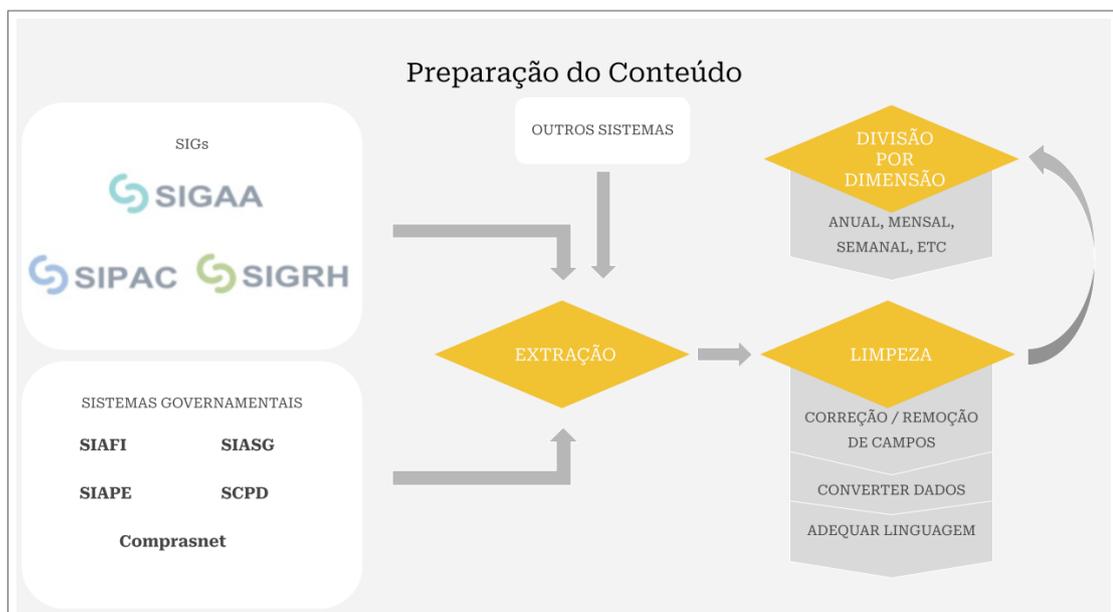
O processo de Extraction-Transformation-Loading (ETL) pode ser definido como uma sequência de etapas com o objetivo de realizar a extração, transformação e limpeza e carga de dados. Os dados podem vir de uma ou mais fontes e recebem tratamento antes de serem inseridas em outra base de dados, como um *data warehouse* (VASSILIADIS, 2009).

Neste Capítulo, é proposta uma abordagem para ETL de dados de IFES com finalidade de realizar carga em portais de dados abertos. Seu principal diferencial consiste em incluir no processo etapas com orientações para garantir a adesão aos requisitos da legislação e às boas práticas de dados abertos governamentais.

### 7.1 Extração e preparação do conteúdo

As bases de dados a serem publicadas no portal são definidas no PDA de cada instituição. A legislação apresenta algumas normas a serem seguidas no processo de disponibilização dessas informações. Esta Seção reúne as recomendações do GTA e da INDA e as consolida em um guia de boas práticas a serem seguidos, ilustrado na Figura 28.

Figura 28 – Processo de publicação - Preparação



Fonte: Autoria Própria (2023)

A primeira etapa consiste na extração dos dados dos sistemas informatizado, o que pode ser feito via relatório gerado pelo sistema ou acessando diretamente sua base de dados. Pode ser

necessário remover campos que não sejam relevantes, converter dados ou ajustar informações que estejam incorretas no sistema.

Em relação ao conteúdo, uma das principais diretrizes consiste em publicar os dados em seu formato primário. Caso o órgão ou entidade deseje realizar algum tipo de agregação, as informações devem ser divulgadas das duas formas, primária e agregada (BRASIL, 2019b). As únicas exceções são o tratamento e remoção de dados pessoais e sensíveis.

Sempre que possível, os dados devem estar em linguagem simples, que possa ser facilmente compreendida pelo usuário, evitando o uso de nomenclaturas pouco conhecidas ou termos técnicos mais específicos. Caso seja necessário fazer uso desses recursos, incluir um glossário ou vocabulário, como o VCGE (BRASIL, 2019b).

Para conjuntos de dados muito grandes, é recomendado fazer sua divisão em subconjuntos menores para facilitar sua manipulação. Pode se usar a dimensão temporal, por ano ou mês, geográfica, ou outra dimensão que o órgão julgue adequada.

## 7.2 Inclusão dos Metadados

A publicação de um conjunto de dados abrange não só os dados em si, mas também os seus metadados. Metadados são dados sobre os dados, ou seja, que permitem organizar, classificar, relacionar e inferir novas informações sobre um conjunto de dados. (BRASIL, 2022a). Essas informações agregam valor ao *dataset* e facilitam que ele seja encontrado e manipulado.

Por recomendação da legislação, os metadados devem conter no mínimo o nome e descrição do conjunto de dados, etiquetas, nome e e-mail do setor responsável pelos dados, periodicidade de atualização, escopo temporal e assuntos do VCGE (BRASIL, 2019b).

Com exceção do último, todos esses campos existem por padrão no CKAN. Para o vocabulário, a aplicação permite que sejam inseridos campos personalizados, que podem ser usados para informar a categoria do VCGE. Todas essas informações se aplicam a todos os recursos associados a um conjunto de dados

Uma informação que não pode deixar de constar nos metadados é a licença aberta, de forma a garantir a permissão irrestrita de uso dos dados. O CKAN conta com várias licenças pré-cadastradas, incluindo a PDDL, que foi escolhida para este trabalho. Ao selecionar a licença neste campo, a ferramenta adiciona o nome da licença nos metadados e um link para sua descrição e texto completo, atendendo assim os critérios necessários para sua correta aplicação. O resultado pode ser visualizado na Figura 29.

Figura 29 – Aplicação da licença PDDL no CKAN



Fonte: Autoria Própria (2023)

Outro componente essencial dos metadados é o dicionário de dados. O documento, que deve acompanhar todo conjunto de dados, possui as seguintes informações sobre os valores contidos em cada coluna do *dataset*: nome, descrição, tipo do dado e obrigatoriedade.

Caso a base contenha algum dado pessoal, é preciso mencionar a hipótese legal que permite sua divulgação sem consentimento do titular. Essa informação pode ser inserida no dicionário. Uma justificativa para remoção de dados sensíveis também pode constar no documento.

Um exemplo desta aplicação é mostrado na Figura 30. A imagem exibe um exemplo de dicionário de dados com a hipótese legal para divulgação de dados pessoais e indicativo de remoção de colunas com dados sensíveis. A segunda parte, embora não seja obrigatória, é importante em caso de atualização de conjuntos de dados para adequação à LGPD, de forma que o público entenda o motivo da remoção ou alteração na base,

Figura 30 – Justificativa para uso de dado pessoal conforme LGPD

**DADOS ABERTOS UFRPE**  
Dicionário de Dados - Docentes - Relatório detalhado  
Atualizado em: 5 de out. de 2022

Informações pessoais divulgadas em concordância com Artigo 31 da Lei de Acesso à Informação (LAI - Lei nº 12.527) e artigo 7º da Lei Geral de Proteção de Dados Pessoais (LGPD - Lei nº 13.709 de 2018), na hipótese de cumprimento de obrigação legal ou regulatória pelo controlador.

Os campos COR\_RACA e DOCENTE\_DEFICIENCIA foram removidos conforme artigo 11º da Lei Geral de Proteção de Dados Pessoais (LGPD - Lei nº 13.709 de 2018).

Campo	Tipo	Descrição	Conteúdo	Obrigatório
CODIGO_IES	Inteiro	Código da IES	-	Sim
NOME_IES	Texto	Nome da IES	-	Sim
NOME_DOCENTE	Texto	Nome do docente	-	Sim
DATA_NASCIMENTO	Data	Data de nascimento do docente	AAAA/MM/DD	Sim

Fonte: Autoria Própria (2023)

Por fim, a Figura 31 mostra os elementos envolvidos na elaboração dos metadados, separando-os em duas categorias, as pertencentes ao dicionário de dados e as que são inseridas em campos próprios do CKAN. Tal distinção foi utilizada como forma de guiar os responsáveis

pela publicação dos dados acerca de onde inserir cada um dos campos obrigatórios.

Figura 31 – Processo de publicação - Metadados



Fonte: Autoria Própria (2023)

### 7.3 Processo de publicação de dados abertos

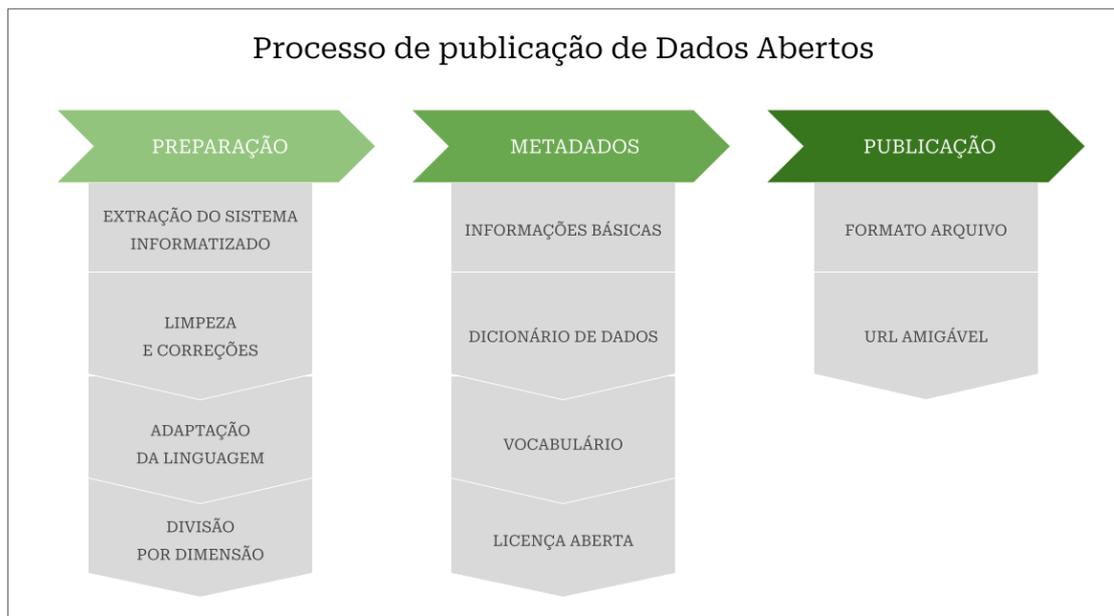
Em relação ao formato, os dados publicados no portal devem estar em formato não proprietário, preferencialmente CSV, JSON, XML, ODS ou RDF. Para a publicação de dados geográficos, a e-PING recomenda Scalable Vector Graphics (SVG) e Geography Markup Language (GML). O formato PDF deve ser usado apenas no dicionário de dados. O uso de compactadores de arquivos é desaconselhado. Caso seja indispensável, deve-se escolher um formato aberto de compactação, como 7Z, GZIP ou ZIP (BRASIL, 2022a).

Outra recomendação importante é a utilização de URLs amigáveis, que permitam ao usuário identificar o conteúdo do conjunto de dados. Todas as palavras devem estar em letras minúsculas, separadas por hífen e sem acentuação. Isso também proporciona uma melhor indexação do conteúdo por motores de busca (BRASIL, 2019b). O CKAN já gera as URLs dos conjuntos de dados nativamente nesse padrão. No entanto, a ferramenta também permite que o autor do *dataset* faça a pré-visualização da URL antes de publicar e realize qualquer

alteração que julgar necessária.

Com base nas considerações apresentadas nesta Seção, foi estabelecido um processo a ser seguido para publicação das bases de dados, que consiste em três fases: preparação, com extração, limpeza e ajustes nos dados; elaboração dos metadados, com cadastro de descrição, vocabulário e dicionário de dados; e publicação no CKAN, com os ajustes finais no formato do arquivo e URL. As etapas desse processo estão ilustradas na Figura 32.

Figura 32 – Processo de publicação de dados abertos



**Fonte:** Autoria Própria (2023)

#### 7.4 Papéis e responsabilidades

Esta Seção detalha os papéis e responsabilidades de cada um dos envolvidos no processo de publicação de dados abertos: os gestores de dados abertos do órgão, a TI e os editores de dados. Tal atribuição é feita pelo PDA das instituições, conforme orientação da Política de Dados Abertos do Poder Executivo Federal (BRASIL, 2016).

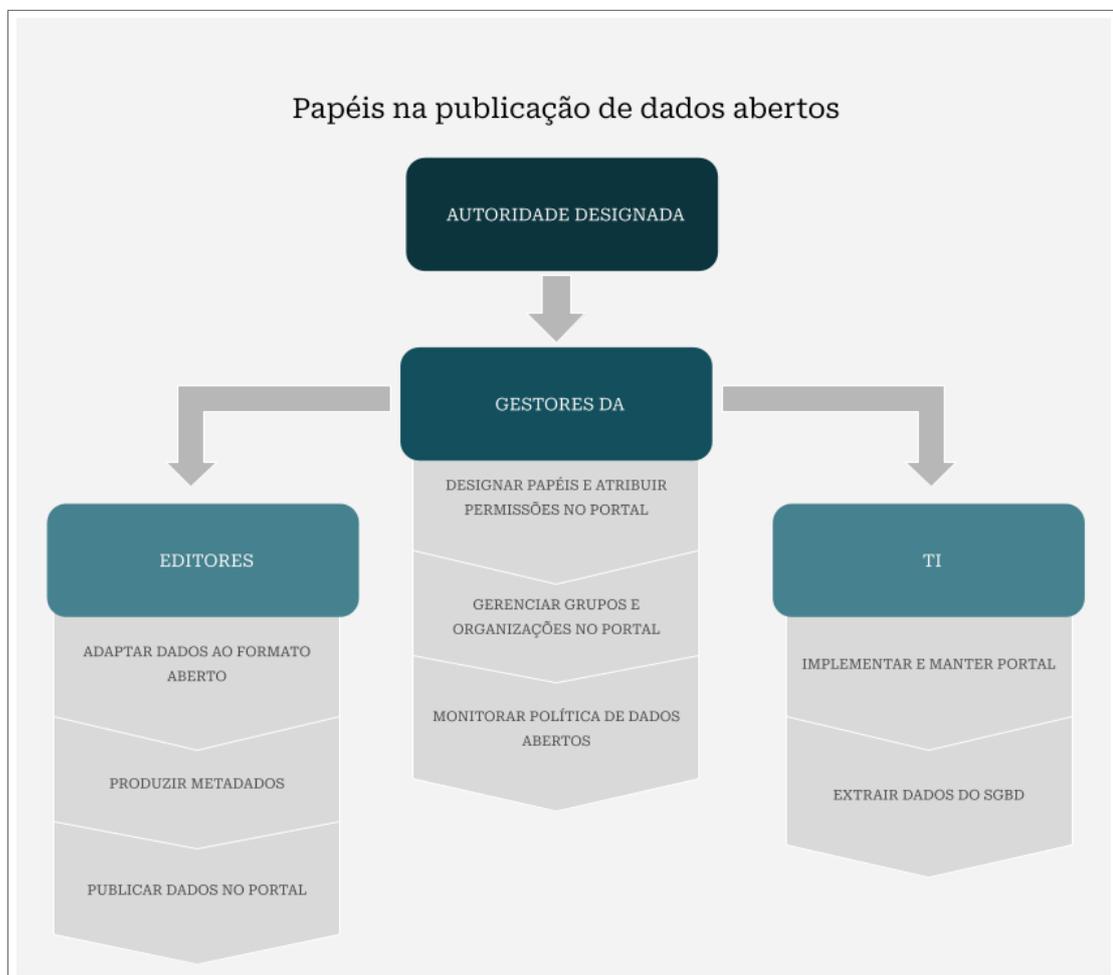
Por determinação da CGINDA e da LAI, cada instituição possui um servidor na função de autoridade designada, que tem como atribuições instituir e monitorar a política de dados abertos daquele órgão (CGINDA, 2017b). Isto inclui a elaboração do PDA, que deve detalhar o papel de cada um dos envolvidos no processo. No caso da UFRPE, foi instituído o CTDA, trabalhando em conjunto com a autoridade designada.

No portal de dados abertos, cabe ao CTDA ou entidade similar a função de fazer o

gerenciamento de permissões de usuários, conforme detalhado no Capítulo 5, e administrar a criação de Grupos e Organizações dentro do CKAN.

Os servidores apontados como responsáveis pela publicação dos dados, por sua vez, têm a responsabilidade de realizar as rotinas de publicação de dados detalhadas neste Capítulo. Já a TI da instituição é responsável por fazer a extração dos dados em casos em que isso não seja possível pela interface do sistema. Também são atribuições da TI a manutenção do portal e gerenciamento de segurança de informação. A Figura 33 mostra o fluxo de responsabilidade de cada um dos papéis envolvidos no processo.

Figura 33 – Papéis na publicação de dados abertos



Fonte: Autoria Própria (2023)

## 7.5 Inserção dos conjuntos de dados no portal

Para por em prática as etapas do processo apresentado nas Seções anteriores, os responsáveis pelos dados devem entrar na página de Conjuntos de Dados e selecionar a opção *Adicionar*

Conjunto de Dados, como mostrado na Figura 34.

Figura 34 – Adicionar conjunto de dados no CKAN



Fonte: Autoria Própria (2023)

Conforme configurado na Seção 5.4 do Capítulo 5, os editores só conseguem criar *datasets* dentro do grupo e organização onde estão inseridos. As opções de *Adicionar Organização* e *Adicionar Grupo* ficam visíveis apenas para os administradores do sistema.

Primeiro são inseridas as informações do Conjunto de Dados, como título, descrição, URL, etiquetas, licença, etc - que se aplicam para todos os recursos dentro do *dataset*. Esses campos são mostrados na Figura 35. Em seguida, deve-se cadastrar o recurso, que pode ser um arquivo de qualquer um dos formatos abertos permitidos. Para inserir um novo recurso em um Conjunto de Dado já existente, o editor deve acessar a opção *Gerenciar* do *dataset*. Nesta tela, mostrada na Figura 36, é possível atualizar ou cadastrar novos recursos.

Figura 35 – Campos do conjunto de dados no CKAN

**1** Criar um conjunto de dados » **2** Adicionar dados

**Título:**

Conjunto

\* URL: dados.ufrpe.br:80/dataset/conjunto

**Descrição:**

ex.: Algumas anotações úteis sobre os dados

Você pode usar formatação [Markdown](#) aqui

**Etiquetas:**

ex.: economia, saúde mental, governo

**Licença:**

Please select the license

Definições de licenças e informações adicionais podem ser encontradas em [opendefinition.org](http://opendefinition.org)

Fonte: Autoria Própria (2023)

Figura 36 – Gerenciar conjunto de dados no CKAN

**Matriculados nas Turmas de Graduação - 2021.2**  
Quantitativos de matrículas nas turmas de graduação no período.

**Matriculados nas Turmas de Graduação - 2021.1**  
Quantitativos de matrículas nas turmas de graduação no período.

**Dicionário de Dados - Matriculados nas turmas...**  
Descrição das colunas da base matriculados nas turmas de graduação.

Fonte: Autoria Própria (2023)

## 8 CONCLUSÕES

Esta pesquisa teve como objetivo principal estudar a implementação de um portal de dados abertos e definir as melhores práticas para publicação das informações definidas pelo PDA das IFES. Através do estudo da literatura e da legislação, verificou-se que um dos principais desafios para o sucesso das iniciativas de transparência ativa consiste em compreender os requisitos necessários, tanto na parte técnica quanto no processo de escolha e tratamento dos dados que serão disponibilizados.

Durante a pesquisa, o CKAN se mostrou a opção mais apropriada de gerenciador de conteúdo. No entanto, apesar de a aplicação ser extremamente versátil e completa, em sua versão padrão ela deixa de atender alguns dos requisitos necessários, como a disponibilização de API e questões de acessibilidade da e-MAG.

Tais limitações foram contornadas através de ajustes nas configurações padrão da aplicação, instalação de *plugins* - nativos ou não - e ajustes no código para melhorar a usabilidade. Apesar de não representar um impedimento para a utilização do CKAN, a necessidade de tais ajustes deve ser levada em conta pelas organizações durante a implantação do portal.

Outro desafio considerável do processo de abertura de dados consiste nos próprios dados. A informação tem que ser vista como um meio, não um fim, e precisa permitir que análises e reusos sejam feitos a partir dela. Para que isso seja possível, é necessário um esforço por parte das organizações para limpar e adequar seus dados.

Conforme visto nos estudos de caso realizados neste trabalho, os setores que utilizam sistemas do Governo Federal com possibilidade de geração de relatórios não podem simplesmente publicá-lo em sua formatação original. É preciso remover informações imprecisas ou irrelevantes, ajustar campos, anonimizar dados pessoais e convertê-los para formatos abertos.

O estudo de caso com as bases de dados de Orçamento e Contratos deixou claro a relevância da cultura organizacional no tratamento das informações e da política de abertura de dados como um todo. A base de Orçamento, por receber esse tratamento de forma rotineira, já se encontrava pronta para publicação. A base de Contratos, por sua vez, precisou passar pelas etapas de limpeza e transformação.

Em contrapartida, o estudo com as bases de Graduação mostrou que os casos em que os dados precisam ser extraídos de SGBDs são ainda mais complexos. Além de todo o trabalho de limpeza e formatação, os órgãos devem se atentar a divulgar os dados em sua forma primária,

permitindo que o público faça suas próprias análises. Isso não impede que as instituições divulguem informações previamente tratadas, com quantitativos ou combinações de duas ou mais bases, desde que não seja em detrimento dos dados primários.

Por fim, tão importante quanto os dados em si são os metadados. Essas informações - como etiquetas, descrição e dicionário de dados - permitem que a sociedade encontre e consiga compreender as informações disponibilizadas. Não menos importante, a licença aberta reafirma para o público o direito de utilizar os dados.

Dessa forma, para garantir que tais orientações sejam seguidas pelos editores, foi definido um guia de boas práticas para o tratamento de dados, englobando desde o processo de extração da informação do sistema informatizado ao seu tratamento e publicação no portal.

O processo de abertura de dados pode ser encarado de duas formas pelas organizações: como um meio de atender as exigências impostas pelo Governo Federal ou na forma de uma ferramenta de participação social e geração de valor para a sociedade.

Com isso, este trabalho entregou sua principal contribuição. Ao abordar os dois aspectos do processo de abertura de dados, técnico e negocial, sempre guiado pela legislação e boas práticas de dados abertos governamentais, o estudo mostra os meios pelos quais as IFES podem obter sucesso em sua iniciativa de abertura, publicando dados de qualidade e, conseqüentemente, incentivando o seu uso pelos dos cidadãos.

## **8.1 Contribuições e Análises dos Objetivos Propostos**

A principal contribuição da pesquisa consiste no seu objetivo geral, alcançado com o estudo da implantação do portal de dados abertos. O processo englobou desde a escolha da aplicação aos ajustes para atender à legislação e às necessidades da instituição.

Quanto aos objetivos específicos, o estudo da literatura serviu como base tanto para a realização da implementação do portal como para a elaboração de um guia de boas práticas para publicação de dados abertos a ser seguido pelos editores das bases. Também pode-se listar como contribuição a integração com o Portal Brasileiro de Dados Abertos.

O estudo de caso, cujas etapas foram realizadas em conjunto com os setores responsáveis pelas informações de Ensino de Graduação, Orçamento e Contratos da instituição, resultou na publicação de 20 conjuntos de dados no portal, distribuídos em quatro grupos.

## 8.2 Limitações da Pesquisa e Trabalhos Futuros

Uma das limitações encontradas no decorrer da pesquisa foi a utilização dos *plugins* do CKAN. Apesar de a ferramenta incentivar o desenvolvimento de extensões, gerando um número considerável de *plugins*, utilizar alguns deles se provou um desafio.

Uma das razões foi falta de compatibilidade com a versão atual do CKAN, o que resultaria em um esforço considerável para adaptar e utilizar a extensão. A falta de documentação também foi um fator determinante para que algumas soluções - como *plugins* adicionais de visualização e para fazer o download de todos os recursos presentes em um *dataset* - não fossem implementadas, assim como a ausência de modularidade em determinados *plugins* que ofereciam várias alterações na ferramenta, mas não permitiam que tais mudanças fossem selecionadas individualmente.

Com base na pesquisa, a autora destaca alguns trabalhos futuros que podem ser realizados de forma a contribuir para a temática:

1. Analisar e propor formas de automatizar a extração dos dados dos SIGs.
2. Estudar estratégia para automatizar as etapas de limpeza e transformação dos arquivos a serem publicados no portal.
3. Estudar melhorias nas opções de acessibilidade do portal.
4. Promover estratégia para abertura de dados de setores pouco - ou não - informatizados.
5. Promover a criação de reúsos para as bases de dados publicadas no portal, como aplicativos e soluções de *software*, incentivando a participação da comunidade acadêmica.

## REFERÊNCIAS

- ARRUDA, W. S. d. Dados abertos governamentais : uma proposta de classificação e estruturação para abertura dos dados de ifes. 2019. Disponível em: <<http://www.tede2.ufrpe.br:8080/tede2/handle/tede2/8238>>. Acesso em: 24 ago. 2022.
- AVILA, T. J. T. Uma proposta de modelo de processo para publicação de dados abertos conectados governamentais. 2015. Disponível em: <<https://www.repositorio.ufal.br/handle/riufal/2209>>. Acesso em: 09 dez. 2022.
- BACHTIAR, A.; SUHARDI; MUHAMAD, W. Literature review of open government data. *2020 International Conference on Information Technology Systems and Innovation (ICITSI)*, p. 329–334, 2020.
- BRASIL. *Constituição*. 1988. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/Constituicao/Constituicao.htm](https://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm)>. Acesso em: 19 out. 2022.
- BRASIL. *e-MAG - Modelo de Acessibilidade em Governo Eletrônico*. 2005. Disponível em: <<https://emag.governoeletronico.gov.br/>>. Acesso em: 19 out. 2022.
- BRASIL. *e-PING - Padrões de Interoperabilidade de Governo Eletrônico*. 2005. Disponível em: <<https://eping.governoeletronico.gov.br/>>. Acesso em: 19 out. 2022.
- BRASIL. *Lei de Acesso à Informação*. 2011. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm)>. Acesso em: 19 out. 2022.
- BRASIL. *Decreto nº 8777*. 2016. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2016/decreto/d8777.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/decreto/d8777.htm)>. Acesso em: 19 out. 2022.
- BRASIL. *Lei Geral de Proteção de Dados Pessoais*. 2018. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm)>. Acesso em: 19 out. 2022.
- BRASIL. *Decreto nº 9903*. 2019. Disponível em: <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2019-2022/2019/decreto/D9903.htm](https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D9903.htm)>. Acesso em: 19 out. 2022.
- BRASIL. Guia de transparência ativa (gta) para os órgãos e entidades do poder executivo federal. *Controladoria Geral da União (CGU)*, 2019. Disponível em: <<https://repositorio.cgu.gov.br/handle/1/46643>>. Acesso em: 19 out. 2022.
- BRASIL. Guia de boas práticas: Lei geral de proteção de dados (lgpd). *Controladoria Geral da União (CGU)*, 2020. Disponível em: <<https://repositorio.cgu.gov.br/handle/1/44262>>. Acesso em: 19 out. 2022.
- BRASIL. Modelo de referência para publicação de dados abertos. *Controladoria Geral da União (CGU)*, 2020. Disponível em: <<https://repositorio.cgu.gov.br/handle/1/46701>>. Acesso em: 07 abr. 2023.
- BRASIL. *Cartilha de Publicação de Dados Abertos*. 2022. Disponível em: <<https://cgugovbr.sharepoint.com/:b:/s/ou-stpc-dtc-cgat/Eaz3GlnxKxxBmofpquodaQkByKXPN7FkC473qtbylEilhw?e=Edmuw1>>. Acesso em: 14 nov. 2022.

BRASIL. *Painel de Monitoramento de Dados Abertos*. 2022. Disponível em: <<http://paineis.cgu.gov.br/dadosabertos/index.htm>>. Acesso em: 26 dez. 2022.

CAMPELO, L. R. R. R.; NETO, V. C. B. Comparando softwares gratuitos para criação de repositórios de dados abertos. *Ciência Da Informação*, 48, 2020. Disponível em: <<https://revista.ibict.br/ciinf/article/view/5004>>. Acesso em: 19 out. 2022.

CGINDA. *Resolução nº 2, de 24 de março de 2017*. 2017. Disponível em: <<https://www.gov.br/governodigital/pt-br/legislacao/resolucao/ginda22432017pdf.pdf>>. Acesso em: 28 out. 2022.

CGINDA. *Resolução nº 3, de 13 de outubro de 2017*. 2017. Disponível em: <[https://www.in.gov.br/materia/-/asset\\_publisher/Kujrw0TZC2Mb/content/id/19357601/do1-2017-10-17-resolucao-n-3-de-13-de-outubro-de-2017-19357481](https://www.in.gov.br/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/19357601/do1-2017-10-17-resolucao-n-3-de-13-de-outubro-de-2017-19357481)>. Acesso em: 19 out. 2022.

CORREA, A. S.; CORREA, P. L. P.; SILVA, F. S. C. da. Transparency portals versus open government data: An assessment of openness in brazilian municipalities. *Association for Computing Machinery*, p. 178–185, 2014. Disponível em: <<https://doi.org/10.1145/2612733.2612760>>. Acesso em: 07 dez. 2022.

COSTA, L. R.; SHINTAKU, M.; SILVEIRA, L. A.; MACEDO, D. J.; FONSECA, R. M. S. d. Guia do usuário ckan. *Ibict*, 2017. Disponível em: <<http://ridi.ibict.br/handle/123456789/1113>>. Acesso em: 15 abr. 2022.

CRAVEIRO, G. S.; MARTANO, A. M. Abertura e disponibilização de dados abertos governamentais: Estudos de caso. 2014. Disponível em: <[http://wtransdev.inf.puc-rio.br/artigos/artigos\\_WTRANS14/paper8.pdf](http://wtransdev.inf.puc-rio.br/artigos/artigos_WTRANS14/paper8.pdf)>. Acesso em: 26 dez. 2022.

DAVIES, T. Ten building blocks of an open data initiative. 2012. Disponível em: <<http://www.opendataimpacts.net/wp-content/uploads/2012/08/Ten-Building-Blocks-of-an-Open-Data-Initiative.pdf>>. Acesso em: 24 out. 2022.

DOCS.CKAN. *Install From Package*. 2022. Disponível em: <<https://docs.ckan.org/en/2.9/maintaining/installing/install-from-package.html>>. Acesso em: 03 abr. 2022.

DOCS.CKAN. *Plugins*. 2022. Disponível em: <<https://docs.ckan.org/en/2.9/maintaining/data-viewer.html>>. Acesso em: 04 abr. 2022.

DOROBĂȚ, I. C.; POSEA, V. Open data indicator: An accumulative methodology for measuring the quality of open government data. *13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2021.

EAVES, D. The three laws of open data. 2009. Disponível em: <<https://eaves.ca/2009/11/29/three-laws-of-open-data-international-edition/>>. Acesso em: 24 out. 2022.

GCF. *Gestão Orçamentária e Financeira UFRPE e UFAPE*. 2022. Disponível em: <<http://www.gcf.ufrpe.br/node/130>>. Acesso em: 18 dez. 2023.

INEP. *Censo da Educação Superior*. 2022. Disponível em: <<https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/censo-da-educacao-superior>>. Acesso em: 08 nov. 2022.

- JAAKKOLA, H.; MAKINEN, T.; ETELAHO, A. Open data: opportunities and challenges. 2014. Disponível em: <<https://dl.acm.org/doi/10.1145/2659532.2659594>>. Acesso em: 03 nov. 2022.
- JANSSEN, M.; CHARALABIDIS, Y.; ZUIDERWIJK, A. Benefits, adoption barriers and myths of open data and open government. 2012. Disponível em: <<https://doi.org/10.1080/10580530.2012.716740>>. Acesso em: 26 dez. 2022.
- KHAYYAT, M.; BANNISTER, F. Open data licensing: More than meets the eye. 2014. Disponível em: <[https://www.researchgate.net/publication/276274546\\_Open\\_Data\\_Licensing\\_More\\_than\\_meets\\_the\\_eye](https://www.researchgate.net/publication/276274546_Open_Data_Licensing_More_than_meets_the_eye)>. Acesso em: 03 nov. 2022.
- LIMA, M. P. d.; ABDALLA, M. M.; OLIVEIRA, L. G. L. A avaliação da transparência ativa e passiva das universidades públicas federais do Brasil à luz da lei de acesso à informação. *Revista do Serviço Público - RSP*, v. 71, p. 232–263, 2020. Disponível em: <<http://repositorio.enap.gov.br/handle/1/5990>>. Acesso em: 19 out. 2022.
- LORENZON, L. N. Análise comparada entre regulamentações de dados pessoais no Brasil e na União Europeia (Lgpd e gdpr) e seus respectivos instrumentos de enforcement. *Revista do Centro de Excelência Jean Monnet da FGV Direito Rio, Rio de Janeiro*, vol. 1, p. 39–52, 2021. Disponível em: <<https://hml-bibliotecadigital.fgv.br/ojs/index.php/rpdue/article/view/83423>>. Acesso em: 05 abr. 2023.
- MARTINS, P. L. Acesso à informação: Um direito fundamental e instrumental. *Acervo - Revista do Arquivo Nacional*, n. 1, v. 24, p. 233–244, 2012. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/43568>>. Acesso em: 19 out. 2022.
- MOLLOY, J. C. *The Open Knowledge Foundation: Open Data Means Better Science*. 2011. Disponível em: <<https://doi.org/10.1371/journal.pbio.1001195>>. Acesso em: 12 abr. 2023.
- MORADI, M.; KEYVANPOUR, M. *CAPTCHA and its Alternatives: A Review*. 2015. 2135–2156 p. Disponível em: <<https://doi.org/10.1002/sec.1157>>. Acesso em: 18 dez. 2022.
- OKFN. *Open Data Commons Public Domain Dedication and License (PDDL)*. 2022. Disponível em: <<https://opendatacommons.org/licenses/pddl/>>. Acesso em: 24 out. 2022.
- OLIVEIRA, W. Q. d. S.; FONSECA, I. F. d. Fatores de sucesso na abertura de dados: o caso do Banco Central do Brasil. *Revista Do Serviço Público*, 72, p. 724–752, 2021. Disponível em: <<https://revista.enap.gov.br/index.php/RSP/article/view/4535>>. Acesso em: 15 dez. 2022.
- OSAGIE, E.; MOHAMMAD, W.; STASIEWICZ, A.; HASSAN, I. A.; PORWOL, L.; OJO, A. State-of-the-art report and evaluation of existing open data platforms. *ROUT-TO-PA*, 2015.
- PDA-UFRPE. *Plano de Dados Abertos da UFRPE*. 2022. Disponível em: <[http://www.acessoainformacao.ufrpe.br/sites/ww2.acessoainformacao.ufrpe.br/files/v06\\_PLANO%20DE%20DADOS%20ABERTOS\\_2022-2024\\_-\\_12.12.2022\\_com%20capas%20Web%20Version.pdf](http://www.acessoainformacao.ufrpe.br/sites/ww2.acessoainformacao.ufrpe.br/files/v06_PLANO%20DE%20DADOS%20ABERTOS_2022-2024_-_12.12.2022_com%20capas%20Web%20Version.pdf)>. Acesso em: 28 out. 2022.
- REIS, A. d. O.; MONTEIRO, D. A. A.; SILVA, A. d. A. P. Qualidade da informação dos dados governamentais abertos: Análise do portal de dados abertos brasileiro. 2020. Disponível em: <[http://wtransdev.inf.puc-rio.br/artigos/artigos\\_WTRANS14/paper8.pdf](http://wtransdev.inf.puc-rio.br/artigos/artigos_WTRANS14/paper8.pdf)>. Acesso em: 26 dez. 2022.

---

SERPRO. *Princípios da LGPD*. 2022. Disponível em: <<https://www.serpro.gov.br/lgpd/menu/tratamento-dos-dados/principios-da-lgpd>>. Acesso em: 21 out. 2022.

SHEA, R.; LIU, J. Understanding the impact of denial of service attacks on virtual machines. p. 1–9, 2012. Acesso em: 07 abr. 2023.

SHINTAKU, M.; SALES, L. *Ciência aberta para editores científicos*. Botucatu, SP: ABEC, 2019. 21-26 p.

SILVA, E. L.; MENEZES, E. M. Metodologia da pesquisa e elaboração de dissertação. UFSC, 2005. Disponível em: <[https://tccbiblio.paginas.ufsc.br/files/2010/09/024\\_Metodologia\\_de\\_pesquisa\\_e\\_elaboracao\\_de\\_teses\\_e\\_dissertacoes1.pdf](https://tccbiblio.paginas.ufsc.br/files/2010/09/024_Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes1.pdf)>. Acesso em: 29 dez. 2022.

SILVA, W. A. de Oliveira e; BRUNI, A. L. Variáveis socioeconômicas determinantes para a transparência pública passiva nos municípios brasileiros. *Revista de Administração Pública [online]*, p. 415–431, 2019. Disponível em: <<https://doi.org/10.1590/0034-761220170383>>. Acesso em: 19 out. 2022.

UBALDI, B. Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance, No. 22*, 2013. Disponível em: <<https://doi.org/10.1787/5k46bj4f03s7-en>>. Acesso em: 1 nov. 2022.

VASSILIADIS, P. A survey of extract–transform–load technology. *International Journal of Data Warehousing and Mining*, p. 1–27, 2009.

## APÊNDICE A – SCRIPTS DE CONFIGURAÇÃO DO CKAN

### Código Fonte 1 – Configurações básicas do CKAN

```

1 ## Front-End Settings
  ckan.site_title = Dados Abertos UFRPE
3 ckan.site_logo = /base/images/ckan-logo.png
  ckan.site_description = Portal de Dados Abertos da UFRPE
5 ckan.favicon = /base/images/ckan.ico

7 ## Internationalisation Settings
  ckan.locale_default = pt_BR
9 ckan.locales_offered = pt_BR en es fr it

11 ## Storage Settings
  ckan.max_resource_size = 20

```

### Código Fonte 2 – Ordenação dos *plugins* no CKAN

```

1 ## Plugins Settings
  ckan.plugins = stats text_view image_view recline_grid_view recline_view
  recline_graph_view pdf_view datastore datapusher
3
  # Define which views should be created by default
5 ckan.views.default_views = image_view text_view recline_view recline_grid_view
  pdf_view

```

### Código Fonte 3 – Desabilitar criação de usuários por aplicação

```

1 ## Authorization Settings
  ckan.auth.create_user_via_api = false
3 ckan.auth.create_user_via_web = false

```

### Código Fonte 4 – Configurações de usuários e permissões

```

## Authorization Settings
2 ckan.auth.anon_create_dataset = false
  ckan.auth.create_unowned_dataset = false
4 ckan.auth.create_dataset_if_not_in_organization = false
  ckan.auth.user_create_groups = false
6 ckan.auth.user_create_organizations = false
  ckan.auth.user_delete_groups = false
8 ckan.auth.user_delete_organizations = false

```