



**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA  
DE TRANSPORTES E GESTÃO TERRITORIAL**

Leonardo Reis Tristão

**PROPOSTA DE UMA BASE DE DADOS INTEGRADA  
PARA APOIAR NO PLANEJAMENTO DO SETOR DE  
AVIAÇÃO CIVIL BRASILEIRO**

Florianópolis

2018



Leonardo Reis Tristão

**PROPOSTA DE UMA BASE DE DADOS INTEGRADA  
PARA APOIAR NO PLANEJAMENTO DO SETOR DE  
AVIAÇÃO CIVIL BRASILEIRO**

Dissertação de mestrado submetida ao Programa de Pós-Graduação em Engenharia de Transportes e Gestão Territorial para a obtenção do Grau de Mestre em Engenharia de Transportes e Gestão Territorial.

Orientador: Prof. Dr. Alexandre Hering Coelho

Florianópolis

2018

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Tristão, Leonardo

Proposição de uma Base de Dados Espacial para Apoiar o Planejamento do Setor de Aviação Civil Brasileiro / Leonardo Tristão ; orientador, Alexandre Hering Coelho, 2018.  
166 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia de Transportes e Gestão Territorial, Florianópolis, 2018.

Inclui referências.

1. Engenharia de Transportes e Gestão Territorial. 2. Planejamento da Aviação Civil Brasileira. 3. Base de dados. 4. Sistema computacional. I. Hering Coelho, Alexandre . II. Universidade Federal de Santa Catarina. Programa de Pós Graduação em Engenharia de Transportes e Gestão Territorial. III. Título.

Leonardo Reis Tristão

**PROPOSTA DE UMA BASE DE DADOS INTEGRADA  
PARA APOIAR NO PLANEJAMENTO DO SETOR DE  
AVIAÇÃO CIVIL BRASILEIRO**

Esta Dissertação de mestrado foi julgada aprovada para a obtenção do Título de “Mestre em Engenharia de Transportes e Gestão Territorial”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Transportes e Gestão Territorial.

Florianópolis, 01 de junho 2018.

---

Prof. Dr. Norberto Hochheim  
Coordenador do Curso

**Banca Examinadora:**

---

Prof. Dr. Alexandre Hering Coelho  
Orientador  
Universidade Federal de Santa Catarina

---

Prof. Dr. Francisco Henrique de Oliveira  
Universidade Federal de Santa Catarina

---

Prof. Dr. Luis Alberto Gomez  
Universidade Federal de Santa Catarina

---

Prof. Dr. Edésio Elias Lopes  
Universidade Federal de Santa Catarina



Este trabalho é dedicado aos meus queridos pais, minha filha amada e minha esposa.



## AGRADECIMENTOS

Aos meus pais Aurino e Margarete, por todo o amor, carinho, dedicação que sempre tiveram comigo, por nunca terem medido esforços para me oferecer o melhor e por acreditar que meus sonhos pudessem virar realidade.

Ao meu eterno amor, minha filha Carol, por todo o amor, por entender minha ausência (mesmo sendo tão pequena), por me dar energia suficiente para eu encarar os desafios da vida e por me ensinar a cada dia algo novo.

Ao meu porto seguro, minha paixão, minha amiga, minha conselheira e esposa, Karina, por todo o amor, zelo e incentivo para encarar cada dia como se não houvesse o amanhã, por embarcar com tudo em todas as minhas loucuras e por estar sempre ao meu lado.

Ao meu irmão, Francisco, pela amizade, por sua fé, pelo companherismo e apoio nesta caminhada.

Ao meu orientador, Alexandre, por sua dedicação, compreensão, amizade e estímulo para que eu pudesse realizar este trabalho.

Aos meus amigos, não citarei nomes para não esquecer de ninguém, por incentivar e torcer por mim em cada meta a ser cumprida.

Ao LabTrans/UFSC, por favorecer a oportunidade de crescimento profissional e pessoal, além de permitir e apoiar a realização do mestrado.



*A persistência é o menor caminho do êxito.  
(Charles Chaplin)*



## RESUMO

A execução de estudos de planejamento no âmbito do transporte aéreo civil é guiada por uma massa de dados consistente, confiável e histórica. Embora haja acessibilidade a esses dados em *sites* e portais governamentais, a sua estrutura, organização, seus padrões e sua documentação são insipientes. Este trabalho consiste em um estudo para o desenvolvimento de uma base de dados integrada, cujo objetivo é apoiar tarefas relacionadas ao planejamento do transporte aéreo brasileiro, focando na utilização de dados públicos. A construção da base de dados gerada foi guiada pelo método proposto. A partir disso, foi desenvolvido um protótipo de um sistema computacional, denominado SIDDA, que permite a visualização dos dados que compõem a base de dados por meio de mapas e indicadores de desempenho. Durante a execução deste trabalho, verificou-se quais conjuntos de dados estão disponíveis de forma pública para o planejamento do transporte aéreo civil, bem como suas fontes. Após a coleta dos dados disponíveis, realizou-se a análise quanto à sua estrutura e ao seu conteúdo, e o resultado foi utilizado como referencial para a modelagem e o subsequente desenvolvimento de uma base de dados integrada. A partir desse cenário, identificou-se a necessidade de desenvolvimento de um protótipo de sistema SIG que permitisse atualizar, consultar e visualizar os dados contidos na base, a fim de não restringir o uso desta a especialistas em sistemas de gerenciamento de banco de dados. Este trabalho busca, portanto, contribuir tanto com a academia quanto com iniciativas governamentais que estejam baseadas na construção, na utilização e na manutenção de bases de dados integradas para o planejamento de transporte aéreo civil.

**Palavras-chave:** Planejamento da Aviação Civil Brasileira. Base de dados integrada. Protótipo de sistema computacional.



## ABSTRACT

The execution of planning studies in the field of civil air transport is guided by consistent, reliable and historical data. While there is accessibility to such data on government websites and portals, its structure, organization, standards, and documentation are insipid. This paper is a study for the development of an integrated database, which aims to support tasks related to brazilian air transport planning, focusing on the use of public data. The construction of the generated database was guided by the proposed method. Then, a computational system prototype, called SIDDA, that allows the visualization of the data that compose the database was developed by maps and performance indicators. During the execution of this work, it was verified which data sets are available for civil air transport planning, as well as its sources. After collecting the available data, the analysis was carried out in terms of its structure and content, and the result was used as a reference for the modeling and subsequent development of a database. In this scenario, it was identified the need to develop a GIS tool prototype that would allow updating, consulting and visualizing the data contained on such database, so as not restricting its use to specialists in database management systems. This work seeks to contribute both to the academy and to governmental initiatives that are based on the construction, use and maintenance of database integrated for civil air transport planning.

**Keywords:** Planning of the Brazilian Civil Aviation. Database integrated. Computational system prototype.



## LISTA DE FIGURAS

Figura 1	Mapa e indicadores temáticos de aeródromos por movimentação. ....	8
Figura 2	Estrutura de uma base de dados. ....	10
Figura 3	Interação entre soluções de base de dados. ....	12
Figura 4	A arquitetura convencional do ETL. ....	14
Figura 5	Ciclo de vida do <i>data warehouse</i> . ....	21
Figura 6	<i>Data mart</i> . ....	22
Figura 7	Etapas do modelo de referência CRISP-DM. ....	26
Figura 8	Estrutura de dados proposta pela INDE para complexo aeroportuário. ....	66
Figura 9	Cenário de aplicação do método para construção de bases de dados. ....	72
Figura 10	Fluxo de atividades do método para construção de bases de dados. ....	79
Figura 11	Modelo estrutural do banco de dados para municípios. .	87
Figura 12	Modelo estrutural do banco de dados para aeródromos. .	89
Figura 13	Modelo estrutural do banco de dados para Horário de transporte. ....	91
Figura 14	Modelo estrutural do banco de dados para movimentação. .	94
Figura 15	Modelo estrutural do banco de dados para voo regular ativo. ....	96
Figura 16	Modelo estrutural do banco de dados para o <i>schema</i> dbo. .	98
Figura 17	Modelo para o <i>schema</i> metadado. ....	99
Figura 18	Comparativo dos totais para indicadores anuais do IBGE. .	102
Figura 19	Validação de dados cadastrais de aeródromos. ....	103
Figura 20	Validação dos dados de pistas. ....	104
Figura 21	Módulos do protótipo e dados utilizados. ....	107
Figura 22	Interface principal do protótipo. ....	108
Figura 23	Seleção de rotas por aeródromo. ....	109
Figura 24	Visualização das principais rotas do aeródromo (linhas de desejo). ....	111
Figura 25	Criação de mapa e indicadores temáticos de aeródromos por movimentação. ....	112

Figura 26	Mapa e indicadores temáticos de aeródromos por movimentação.....	113
Figura 27	Criação de mapa e indicadores temáticos para dados do IBGE.....	115
Figura 28	Mapa e indicadores temáticos por municípios para dados do IBGE.....	116
Figura 29	Mapa e indicadores temáticos por estados para dados do IBGE.....	117
Figura 30	Gráficos para dados de movimentação aeroportuária - passageiros.....	119
Figura 31	Gráficos para dados do IBGE.....	120
Figura 32	Gráficos para dados VRA.....	122
Figura 33	Visualização de dados completos.....	125
Figura 34	Atualização de Horário de transporte.....	128
Figura 35	Inclusão de novos dados.....	129
Figura 36	Processo de vinculação de novos dados.....	131
Figura 37	Visualização dos novos dados.....	132
Figura 38	Criação de tema com novos dados de aeródromos.....	133
Figura 39	Tema com novos dados de aeródromos.....	145
Figura 40	SBFL - Movimentação de aeronaves.....	148
Figura 41	SBFL - Movimentação de passageiros.....	149
Figura 42	SBFL - Movimentação de carga aérea.....	149
Figura 43	SBFL - Movimentação de mala postal.....	150
Figura 44	SBFL - Projeção da movimentação de passageiros.....	152
Figura 45	SBFL - Principais fluxos (linhas de desejo).....	153

## LISTA DE TABELAS

Tabela 1	Resumo dos dados identificados.....	70
Tabela 2	Registro de fontes primárias de dados.....	82
Tabela 3	Registro da coleta dos dados.....	134
Tabela 4	Análise da estrutura dos dados de municípios.....	135
Tabela 5	Dados compilados de unidade da federação.....	136
Tabela 6	Amostra dos dados compilados de municípios.....	137
Tabela 7	Dados compilados de indicadores dos municípios (2010).	138
Tabela 8	Análise da estrutura dos dados de aeródromos.....	139
Tabela 9	Amostra dos dados compilados do cadastro de aeródromos.	140
Tabela 10	Amostra do dados compilados de pistas dos aeródromos.	140
Tabela 11	Análise da estrutura dos dados de HOTRAN.....	141
Tabela 12	Análise da estrutura dos dados de movimentação.....	142
Tabela 13	Amostra dos dados compilados de companhias aéreas. .	143
Tabela 14	Análise da estrutura dos dados de VRA.....	143
Tabela 15	Exemplo de metadados de tabelas.....	144
Tabela 16	Comparativo dos dados de horário de transporte.....	144
Tabela 17	Florianópolis - Indicadores socioeconômicos.....	151



## LISTA DE ABREVIATURAS E SIGLAS

ABDI	Agência Brasileira de Desenvolvimento Industrial
AISWEB	Serviço de Informação Aeronáutica
ANAC	Agência Nacional de Aviação Civil
API	Application Programming Interface
ASK	Available Seat Kilometers
ATK	Available Tonne Kilometres
AZUL	AZUL Linhas Aéreas
BAV	Boletins de Alteração de Voo
CSV	Comma Separated Values
DECEA	Departamento de Controle do Espaço Aéreo
EGB	Espaço Geográfico Brasileiro
ETL	Extract, Transform and Load
FOB	Free On Board
FTP	File Transfer Protocol
GB	Gigabyte
GOL	Gol Linhas Aéreas Inteligentes S.A
HOTRAN	Horário de Transporte
HTML	Hyper Text Markup Language
IBGE	Instituto Brasileiro de Geografia e Estatística
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
IDE	Infraestrutura de Dados Espaciais
IFR	Instrument Flight Rules
IG	Informação Geográfica
INDE	Infraestrutura Nacional de Dados Espaciais
INFRAERO	Empresa Brasileira de Infraestrutura Aeroportuária
INMETRO	Instituto Nacional de Metrologia, Qualidade e Tecnologia
INPI	Instituto Nacional da Propriedade Industrial
KDD	Knowledge Discovery in Database
KG	Quilograma
KM	Quilômetro
LAP	Linhas Aéreas Paulistas

MB	Megabyte
MDIC	Ministério da Indústria, Comércio Exterior e Serviços
METAR	Informe Meteorológico Regular de Aeródromo
MUNIC	Pesquisa de Informações Básicas Municipais
NAB	Navegação Aérea Brasileira
NCM	Nomenclatura Comum do Mercosul
NOTRAM	Notice to Airmen
SAC/MTPA	Secretaria Nacional de Aviação Civil do Ministério dos Transportes, Portos e Aviação Civil
SBGR	Código do aeródromo de Guarulhos
SBFL	Código do aeródromo Herclílio Luz
SGBD	Sistema de Gerenciamento de Banco de Dados
SIDDA	Sistema de Integração e Disponibilização de Dados Aeroportuários
SIG	Sistema de Informações Geográficas
SISCOMEX	Sistema Integrado de Comércio Exterior
SISCEAB	Sistema de Controle do Espaço Aéreo Brasileiro
SUFRAMA	Superintendência da Zona Franca de Manaus
OACI	Organização da Aviação Civil Internacional
PAN	Plano Aeroviário Nacional
PCN	Pavement Classification Number
PDAR	Plano de Desenvolvimento da Aviação Regional
PIB	Produto Interno Bruto
PIL	Programa de Investimentos em Logística: Aeroportos
PNDIA	Plano Nacional de Desenvolvimento da Infraestrutura Aeronáutica Civil
PROFAA	Programa Federal de Auxílio a Aeroportos
ROTAER	Publicação Auxiliar de Rotas Aéreas
RPK	Revenue Passenger Kilometres
RTK	Revenue Tonne Kilometres
TABA	Transportes Aéreos Bandeirantes
TABA	Transportes Aéreos da Bacia Amazônica
TAM	Táxi Aéreo Marília
TAM	Transportes Aéreos Regionais S/A

TRIP	TRIP Linhas Aéreas
URL	Uniform Resource Locator
VARIG	Viação Aérea Rio-Grandense
VASD	Viação Aérea Santos Dumont
VASP	Viação Aérea São Paulo
VFR	Visual Flight Rules
VRA	Voo Regular Ativo
XLS	Extensão de planilha do Microsoft Excel
XML	eXtensible Markup Language
ZIP	Extensão de arquivo compactado



## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>1</b>
1.1	OBJETIVOS .....	3
1.1.1	Objetivo geral .....	3
1.1.2	Objetivos específicos .....	4
1.1.3	Delimitação de escopo .....	4
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	<b>5</b>
2.1	CONSTRUÇÃO DE BASE DE DADOS PARA O PLANEJAMENTO DA AVIAÇÃO CIVIL .....	5
2.1.1	Processo ETL .....	13
2.1.2	<i>Data Warehouse</i> .....	16
2.1.3	<i>Data Mart</i> .....	22
2.1.4	<i>Data Mining</i> .....	24
2.2	DADOS ABERTOS GOVERNAMENTAIS .....	28
2.2.1	Instituto Brasileiro de Geografia e Estatística .....	30
2.2.2	Agência Nacional de Aviação Civil .....	31
2.2.3	Empresa Brasileira de Infraestrutura Aeroportuária .....	42
2.2.4	Departamento de Controle do Espaço Aéreo .....	46
2.2.5	Secretaria de Aviação Civil do Ministério dos Transportes, Portos e Aviação Civil .....	50
2.2.6	Ministério da Indústria, Comércio Exterior e Serviços .....	58
2.2.7	Resumo .....	61
2.3	INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS .....	61
<b>3</b>	<b>MÉTODO PARA CONSTRUÇÃO DA BASE DE DADOS</b> .....	<b>71</b>
<b>4</b>	<b>APLICAÇÃO DO MÉTODO PARA CONSTRUÇÃO DE BASE DE DADOS</b> .....	<b>81</b>
4.1	IDENTIFICAÇÃO CONJUNTO DE DADOS E FONTES DE DADOS .....	81
4.2	COLETA DOS DADOS .....	82
4.2.1	Análise dos dados coletados .....	84
4.2.2	Dados de estados e municípios .....	84
4.2.3	Dados de cadastro de aeródromos .....	86
4.2.4	Dados de horário de transporte .....	90
4.2.5	Dados de movimentação aeroportuária .....	92
4.2.6	Dados de voo regular ativo .....	93
4.3	CRIAÇÃO DA ESTRUTURA DE BANCO DE DADOS .....	95

4.4	REGISTRO DE METADADOS .....	97
4.5	INSERÇÃO DOS REGISTROS .....	100
4.6	VALIDAÇÃO DOS DADOS .....	101
4.7	DESENVOLVIMENTO DE SISTEMA PARA USO DOS DADOS .....	105
<b>4.7.1</b>	<b>Sistema de Integração e Disponibilização de Dados Aeroportuários .....</b>	<b>106</b>
4.7.1.1	Principais rotas entre aeródromos .....	107
4.7.1.2	Mapas e indicadores temáticos .....	110
4.7.1.3	Gráficos e infográficos .....	115
4.7.1.4	Dados completos .....	121
4.7.1.5	Atualização e inclusão de dados .....	124
<b>5</b>	<b>EXPERIMENTO DE CARACTERIZAÇÃO DE UM AEROPORTO .....</b>	<b>147</b>
<b>6</b>	<b>CONCLUSÕES E SUGESTÕES PARA TRABA- LHOS FUTUROS.....</b>	<b>155</b>
6.1	CONCLUSÕES .....	155
6.2	RECOMENDAÇÕES PARA TRABALHOS FUTUROS ..	160
	<b>REFERÊNCIAS .....</b>	<b>161</b>

## 1 INTRODUÇÃO

Com o advento dos sistemas de transportes, os mercados consumidores passaram a configurar-se como grandes potenciais de exploração. Conforme Moraes (2005), o progresso das civilizações está associado com a eficácia e a eficiência com que as mercadorias e pessoas se deslocam, isso por meio de seus meios de transportes. Devido à necessidade de subsistência do homem, o volume de bens, mercadorias e pessoas em deslocamento continua crescendo em um ritmo acelerado.

Tadeu (2010) destaca que, para o Brasil – assim como para o resto do mundo –, os meios de transportes são subsídio para o crescimento econômico. No entanto, a infraestrutura de transportes brasileira tem uma realidade diferente daquela de diversos países com o mesmo potencial econômico. Para o autor, tal infraestrutura, mesmo não sendo propícia, é fortemente estimulada por um crescimento favorável, e destaca que, para lidar com as limitações existentes, há a necessidade da adoção de opções tecnológicas disponíveis para o setor.

Desde a realização do primeiro voo comercial, em 1927, o setor de aviação civil brasileiro vem crescendo e apresentando-se como um mercado promissor (MCKINSEY&COMPANY, 2010). Entre os anos de 2007 e 2016, o setor registrou uma taxa média de crescimento ao ano de 10,20% na movimentação de passageiros e 5,62% na movimentação de aeronaves. Isso representa que, no ano de 2016, os aeroportos brasileiros movimentaram 201,22 milhões de passageiros para um cenário de mais de 1,78 milhões de pousos e decolagens (ANAC, 2017b).

Diversas frentes estão sendo desenvolvidas para apoiar o planejamento de transportes aéreo do país. Conforme descrito por SAC (2014), existem variadas iniciativas governamentais alinhadas ao planejamento do setor, como é o caso do Plano Aeroviário Nacional (PAN), do Programa Federal de Auxílio a Aeroportos (PROFAA), do Plano Nacional de Desenvolvimento da Infraestrutura Aeronáutica Civil (PNDIA), do Plano de Desenvolvimento da Aviação Regional (PDAR) e do Programa de Investimentos em Logística: Aeroportos (PIL).

Conforme descrito por Simões (2003), o conjunto de dados a ser considerado no planejamento deve envolver vários aspectos da área de aviação civil, como movimentação de aeronaves, passageiros, carga aérea, mala postal e infraestrutura atual, bem como aspectos econômicos da região onde o aeródromo está situado. Segundo o autor, é comum que, durante a seleção dos conjuntos de dados para tais estudos, sejam identificados tanto dados alfanuméricos quanto geográficos, pois

quando se trata de dados relacionados ao setor, é evidente que eles estejam relacionados a aeródromos ou municípios, podendo conter informações geoespaciais.

Em iniciativas de planejamento do setor de aviação civil, é necessária uma massa de dados consistente, confiável e histórica. Segundo Tadeu (2010), os dados existem e estão acessíveis, porém, na maioria dos casos, não estão disponíveis de forma organizada, padronizada, consolidada, correlacionada, estruturada e acompanhados de uma documentação sobre sua estrutura. Nas etapas iniciais de um projeto de planejamento, é custosa a construção de uma base de dados para o fornecimento de informações voltadas à sua execução, isso porque a base necessária normalmente envolve diferentes fontes e formatos de dados.

Muitos dos dados demandados para o planejamento da aviação civil são disponibilizados por instituições e órgãos competentes de forma pública, atendendo, desse modo, à Lei de Acesso à Informação<sup>1</sup>; nesse contexto, podem ser citados os seguintes:

- Agência Nacional de Aviação Civil (ANAC);
- Departamento de Controle do Espaço Aéreo (DECEA);
- Empresa Brasileira de Infraestrutura Aeroportuária (INFRAERO);
- Instituto Brasileiro de Geografia e Estatística (IBGE);
- Ministério do Desenvolvimento, Indústria e Comércio Exterior (MDIC); e
- Secretaria Nacional de Aviação Civil do Ministério dos Transportes, Portos e Aviação Civil (SAC/MTPA).

A título de exemplo, cita-se a base de dados HOTRAN (sigla para "Horário de Transporte"), publicada pela Agência Nacional de Aviação Civil (ANAC) com a finalidade de formalizar as concessões para a exploração de linhas aéreas regulares internacionais e domésticas de passageiros, carga e rede postal pelas empresas de transporte aéreo, com os respectivos horários, números de voos, frequências, tipos de aeronaves e oferta de assentos (ANAC, 2017c). Para a área de estudo da aviação civil brasileira, tal conjunto de dados pode ser utilizado para o planejamento de demanda de voos, além do estabelecimento de matriz origem/destino (rotas aéreas).

---

<sup>1</sup>BRASIL. Lei n. 12.527, de 18 de novembro de 2011.

As bases de dados estatísticas da ANAC apresentam um conjunto de informações sobre a movimentação de passageiros, cargas, aeronaves e mala postal referente aos aeródromos nacionais, trazendo informações para origens e destinos de voos doméstico e internacionais a partir do ano de 2000 (ANAC, 2017b).

Na listagem pública de aeródromos da ANAC, estão inseridas diversas informações cadastrais acerca dos aeródromos nacionais públicos e privados (ANAC, 2017a). Para complementar os dados dos aeródromos, existem os dados políticos referentes a municípios via Instituto Brasileiro de Geografia e Estatística (IBGE), os quais podem agregar informações ao cadastro de aeródromos, tais como PIB e população (IBGE, 2017a).

Segundo Filho (2012), é possível confirmar que, de modo geral, os dados existem e são acessíveis, conforme citado. Uma base de dados, quando construída seguindo métodos, normas e processos, pode ser reaproveitada em outros estudos, o que acarretaria na redução de custos inerentes ao levantamento de informações (QUEIROZ, 2014).

Uma abordagem que apoie a construção de bases de dados voltadas ao fornecimento de informações para o planejamento da aviação civil pode ser vista como um mecanismo de coleta, integração e consolidação de dados, permitindo a realização do cruzamento, agrupamento, atualização e reutilização das informações em novos estudos de planejamento no setor. Dessa maneira, a questão de pesquisa proposta é: quais elementos são necessários para a construção de uma bases de dados integrada para o planejamento do setor de aviação civil, de maneira a otimizar a coleta, a integração e a consolidação dos dados, permitindo a atualização e a reutilização deles?

## 1.1 OBJETIVOS

Esta seção apresenta os objetivos e as delimitações do trabalho.

### 1.1.1 Objetivo geral

Propor uma bases de dados integrada, com a finalidade de apoiar tarefas relacionadas ao planejamento do transporte aéreo civil brasileiro, com foco em dados públicos e em formato aberto. .

### 1.1.2 Objetivos específicos

Foram traçados os seguintes objetivos específicos:

1. Verificar quais dados estão disponíveis de forma aberta e sistemática para apoiar o planejamento do transporte aéreo civil brasileiro;
2. Coletar dados públicos disponíveis na internet de caracterização e operação acerca dos aeródromos públicos e privados brasileiros;
3. Identificar as atividades necessárias para construção de uma base de dados integrada para apoiar no planejamento do transporte aéreo civil brasileiro;
4. Desenvolver um protótipo de sistema voltado à disponibilização de dados e indicadores, baseadas na base de dados gerada;
5. Apresentar experimentos de uso do protótipo de sistema.

### 1.1.3 Delimitação de escopo

Estabelecido o objetivo geral, entende-se que este trabalho não se concentrará no estabelecimento de processos e cálculos numéricos envolvidos no planejamento e na operação dos aeródromos, também não tem como objetivo listar exaustivamente todos os dados envolvidos no planejamento. Quanto ao sistema a ser desenvolvido, o mesmo terá a característica de protótipo, neste sentido o mesmo não contemplará todas as funcionalidades de um sistema SIG, nem passará por testes avançados a fim de garantir a ausência de erros.

## 2 REFERENCIAL TEÓRICO

A revisão de literatura aborda as temáticas principais desta pesquisa: a primeira refere-se à fundamentação teórica, que trata dos assuntos relacionados à estruturação de base de dados; já a segunda temática objetiva apresentar os dados públicos disponíveis na internet relacionados ao setor de aviação civil brasileira.

### 2.1 CONSTRUÇÃO DE BASE DE DADOS PARA O PLANEJAMENTO DA AVIAÇÃO CIVIL

O Brasil, como todos os países, necessita cada vez mais uma rede de transportes eficiente. Investimentos em infraestrutura são de extrema importância, pois garantem a produtividade dos sistemas de transportes. Segundo Padula (2008), o planejamento permite identificar deficiências do sistema de transportes, auxiliando quanto ao estabelecimento de soluções concretas para contornar tais problemas. Quando executado de forma consistente, o planejamento garante que as condicionantes que circundam o sistema de transporte não resultem em equívocos; caso contrário, erros do passado continuarão a ser reproduzidos, podendo agravar o cenário existente.

O planejamento de transportes é uma combinação de várias ações, com a finalidade de solucionar problemas ocasionados pelos modos operantes, no intuito de atingir metas e objetivos previamente estabelecidos. Para Abreu (2008), durante o planejamento, as decisões ocasionam efeitos e consequências que irão ocorrer ao longo do tempo. Todavia, tais ações podem provocar mudanças no modelo de operação de um sistema de transporte como um todo.

De acordo com Horonjeff et al. (2010), para o setor de aviação civil, um plano é uma representação da situação das instalações aeroportuárias visando satisfazer às necessidades imediatas e futuras de uma área metropolitana, uma região, um estado ou até mesmo um país. Ashford, Mumayiz e Wright (2011) afirmam que, durante um processo de planejamento, devem ser considerados cuidadosamente o papel, a função e a interação dos aeroportos, visando ao sistema de transporte como um todo. Caso contrário, as ações resultantes do planejamento poderão implicar em impactos negativos no desenvolvimento de áreas metropolitanas, regiões, estados ou mesmo de um país.

Para Kazda e Caves (2007), planejar o desenvolvimento de um

aeroporto normalmente é complicado, pois existem diferentes tipos de infraestrutura e níveis tecnológicos nas instalações aeroportuárias. Um exemplo são as instalações voltadas para operações com passageiros, que necessitam de equipamentos diferentes das instalações voltadas para operação com cargas.

Ashford, Mumayiz e Wright (2011) explana que a quantidade de dados necessária para a execução de um planejamento é muito extensa, mesmo que esta seja relacionada a apenas um aeroporto ou a um conjunto de aeroportos delimitado. Comumente, obter todos os dados é uma atividade trabalhosa em termos de tempo e custo; ademais, considerando-se que, em iniciativas desse tipo, os orçamentos são limitados, a alternativa viável é a coleta apenas dos dados mais importantes e que estejam acessíveis.

Conforme Horonjeff et al. (2010), são muitos os temas de interesse quando se pensa na realização do planejamento aeroportuário. Normalmente, tais interesses incluem estudos relacionados às áreas de infraestrutura, situação financeira, tráfego, mercados, economia e meio ambiente. Para Kazda e Caves (2007), os resultados desses estudos são propostas que visam ao desenvolvimento da infraestrutura aeroportuária. Nesse sentido, em Horonjeff et al. (2010), são destacados os planos do sistema aeroportuário e os planos diretores como elementos cruciais para o sistema de transporte aéreo como um todo.

Os planos diretores servem como um guia para o desenvolvimento em um aeroporto. Seu foco está no atendimento a demandas previstas, permitindo maximizar e preservar a capacidade final do aeroporto. Nem sempre é possível desenvolver uma solução dogmática para planos diretores, pois normalmente é necessário buscar alternativas específicas para cada aeroporto. Para que o plano tenha devida importância, ele deve ser tratado como um termo de compromisso para o desenvolvimento do aeroporto (KAZDA; CAVES, 2007).

Para Ashford, Mumayiz e Wright (2011), o plano diretor pode ser desenvolvido para a implantação de um aeroporto novo ou, ainda, como um conjunto de alternativas para a expansão dos aeroportos existentes. Normalmente, os planos diretores abrangem cenários de curto, médio ou longo prazos, visando garantir o desenvolvimento do aeroporto. Tais planos devem ser revistos no mínimo a cada cinco anos, para que suas informações e os seus objetivos sejam atualizados de forma incremental (FDT, 2010).

Horonjeff et al. (2010) descreve um plano de sistema aeroportuário como uma análise em que os aeroportos são tratados como uma rede de transportes. Para o autor, nesse tipo de planejamento, é

possível identificar o papel da aviação de novos aeroportos em relação ao sistema aéreo existente. Ou seja, as atividades de aviação de um aeroporto impactam diretamente nos demais elementos da rede aérea.

Durante o processo de construção dos planos aeroportuários, muitas informações que fogem dos aspectos de transportes devem ser consideradas, tais como planos regionais e estaduais, além dos objetivos nacionais para o modal de transporte, pois as relações entre os elementos do sistema de transportes como um todo afetam diretamente o sistema aeroportuário. Horonjeff et al. (2010) ressalta a importância de analisar aspectos planejados por outras iniciativas do governo, com o intuito de mensuração de seu impacto no sistema aeroportuário.

Para Ashford, Mumayiz e Wright (2011), os dados necessários para o plano do sistema aeroportuário são semelhantes (em estrutura) àqueles utilizados em um plano diretor típico (figura 1); porém, diferem de duas formas. Primeiramente, a granularidade dos dados necessários para a análise de um sistema aeroportuário em relação ao plano diretor está em nível macro aeroporto (não detalhada). Ainda, os tipos de dados e as análises adaptadas para o sistema como um todo são diferentes da análise plano diretor, pois um plano de sistema aeroportuário analisa o sistema como uma rede e o plano diretor avalia apenas um nó em uma rede.

Segundo McKinsey&Company (2010), a visão desejada para o setor de aviação civil brasileira é que ele chegue em seu "pleno potencial", culminando em significativos benefícios sociais, alcançando produtividade e tendo uma melhor utilização dos ativos, bem como garantindo que o sistema de administração aeroportuário seja autossuficiente. Nesse cenário, as companhias aéreas nacionais operariam com alto nível de eficiência e os passageiros poderiam tirar proveito via redução dos preços de passagens aéreas.

É possível identificar iniciativas com foco em desenvolvimento realizadas pelo Governo Federal a fim de apoiar o planejamento de transportes para a aviação civil brasileira ( SAC (2014)), podendo-se destacar:

- **Plano Aeroviário Nacional (PAN):** tem como objetivo adequar a infraestrutura de transporte aéreo existente, coordenando a aplicação dos investimentos necessários no setor.
- **Plano Nacional de Desenvolvimento da Infraestrutura Aeronáutica Civil (PNDIA):** visa ao desenvolvimento integrado da infraestrutura de transporte aéreo do Brasil, considerando tanto o espaço aéreo quanto a infraestrutura existente.

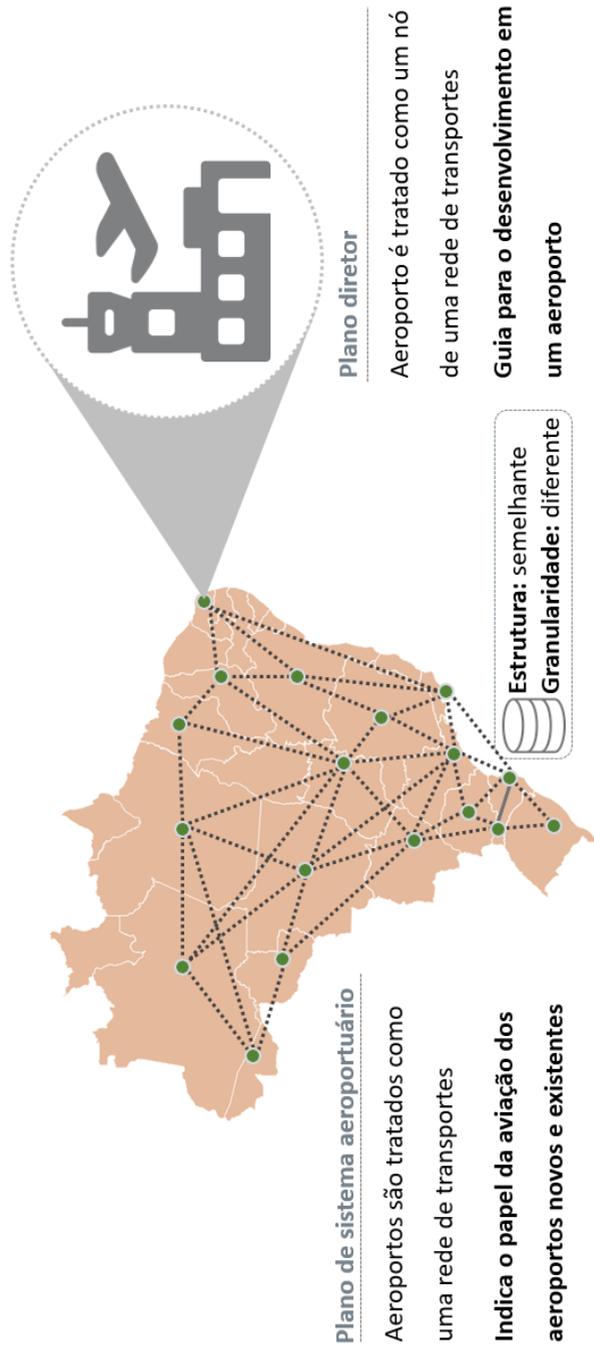


Figura 1 – Mapa e indicadores temáticos de aeródromos por movimentação.

Fonte: Adaptada de Kazda e Caves (2007), Horonjeff et al. (2010), Ashford, Mumayiz e Wright (2011)

- **Plano de Desenvolvimento da Aviação Regional (PDAR):** busca desenvolver e expandir o transporte aéreo para rotas de baixo e médio potencial de tráfego, além da ampliação da oferta para o transporte aéreo de passageiros e de cargas.
- **Programa Federal de Auxílio a Aeroportos (PROFAA):** tem como objetivo a construção, a reforma e a ampliação de infraestrutura aeroportuária de interesse regional e estadual.
- **Programa de Investimentos em Logística: Aeroportos (PIL):** objetiva fortalecer e ampliar a aviação regional por meio de investimentos, evoluir o modelo de regulação para horários de chegadas e partidas em aeroportos que operam no limite da capacidade e autorizar a operação para aeroportos dedicados à aviação geral.
- **Programa de Aviação Civil:** visa estabelecer objetivos que viabilizem a adequação da capacidade e a expansão das infraestruturas aeroportuárias, apoiando o transporte de passageiros e cargas.

Tanto nas iniciativas do Governo Federal (Planos e Programas) como na literatura, é possível identificar uma fase comum a todos: a necessidade de levantamento de dados a respeito de operação, infraestrutura e características socioeconômicas da região sede dos aeródromos. Portanto, é evidente a demanda por informações consolidadas e consistentes para a execução de estudos para o setor de aviação civil.

Para FDT (2010), o desenvolvimento de uma base de dados é normalmente uma das primeiras tarefas a serem realizadas para a execução de estudos no setor de aviação civil, sendo o seu principal objetivo reunir e montar um acervo de informações sobre o setor. A qualidade das informações é fundamental para qualquer projeto de planejamento na área de transportes, seja no setor aéreo, rodoviário, ferroviário ou hidroviário. Em sua essência, os projetos de planejamento, principalmente do setor de aviação civil brasileiro, têm a necessidade de um grande volume de dados. E, para mantê-los estruturados, interligados e correlacionados, garantindo a sua qualidade, é importante que sua organização ocorra em um banco de dados.

Entende-se banco de dados como um acervo de informações a respeito de determinado assunto em meio físico ou digital. No contexto deste trabalho, um banco de dados deve ser visto como um sistema de computador (*software*) que objetiva estruturar, armazenar, atualizar e recuperar dados e informações de forma eficiente.

Para DAMA (2012), um dado pode ser caracterizado como o registro de um evento ou de uma ação que normalmente ocorreu em local e período de tempo determinados. Quando os dados são submetidos a um processo de análise e interpretação, pode-se dizer que foi iniciado um processo de busca de conhecimento. Portanto, informação pode ser caracterizada como conhecimento, que é resultante da análise de um conjunto de dados. Comumente, dados e informações existem de forma conjunta, pois um depende diretamente do outro. Não existe informação sem um conjunto de dados; todavia, dados por si só não agregam conhecimento aos negócios.

Conforme apresentado em Date (2003), em um banco de dados, estes são armazenados em estruturas denominadas tabelas, as quais, por sua vez, são formadas por colunas e linhas (Figura 2). As colunas representam um tipo de dado e as linhas armazenam o registro do dado propriamente dito. Portanto, tabelas são conjuntos de colunas estruturadas voltadas ao armazenamento de determinado tipo de dado. Normalmente, cada coluna comporta somente dados com o mesmo tipo estrutural, isso para garantir eficiência na manipulação e na organização dos dados. Um dado pode ser classificado com base na natureza de seu registro, como em número inteiro, número decimal, texto, booleano, caractere simples ou ainda um dado geoespacial.

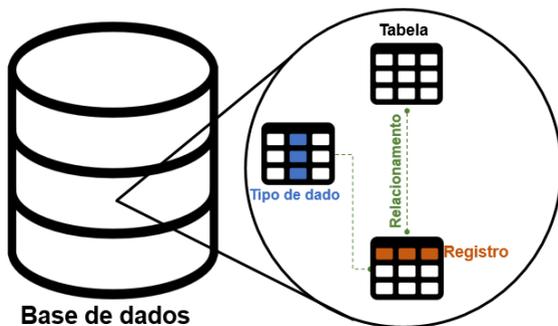


Figura 2 – Estrutura de uma base de dados.

Date (2003) e DAMA (2012) explanam que um banco de dados é formado por informações das mais variadas conjecturas, sendo sua estrutura composta por tabelas para a representação de cada tipo de informação. Para que as informações estejam interligadas e correlacionadas, os bancos de dados permitem a criação de relacionamentos entre tabelas. Relacionamento é a forma de estabelecer que colunas de uma

tabela representam (ou estão referenciando) valores de uma coluna que estão em outra tabela. Ou seja, relacionamentos em bancos de dados são ligações entre colunas de tabelas diferentes, com a finalidade de referenciar, organizar, manter ou especificar os dados.

Durante a construção de uma base de dados, é importante identificar seu principal objetivo e qual tipo de uso é almejado. Existem inúmeras técnicas que apoiam a estruturação de uma base de dados, permitindo o atendimento a objetivos e tipos de uso. Algumas técnicas têm foco na estruturação da base de dados a fim de obter melhor desempenho durante a recuperação de dados; outras buscam a melhor organização possível das informações. Em projetos de banco de dados, a aplicação de técnicas mistas permite que parte da estrutura seja voltada para desempenho e parte para organização. Isso posto, a estratégia adotada para a construção da base de dados guiará o trabalho de modelagem da estrutura de tabelas.

A modelagem pode ser compreendida como um conjunto de atividades com a finalidade de estabelecer a estrutura necessária para o armazenamento de dados. Durante a modelagem, as tabelas, as colunas e os relacionamentos são documentados por meio de diagramas, os quais permitem a visualização gráfica da estrutura de dados até mesmo antes da conclusão da construção da base de dados. Para DAMA (2012), os diagramas apoiam no entendimento da estrutura de tabelas e, por conseguinte, nas manutenções das estruturas da base de dados. Frequentemente, o conjunto de dados que faz parte da composição da base de dados e a finalidade de uso definem o modelo para as tabelas e os relacionamentos a serem estabelecidos. Sendo assim, o resultado da análise dos dados (quanto a possíveis relacionamentos) orienta o processo de modelagem.

Para o setor de transportes aéreo civil, quando a base de dados entra em operação, ela demanda um grande volume de informações. Portanto, é comum a identificação de necessidades de integração com outras bases de dados, com a finalidade de disponibilizar ou agregar novos conhecimentos. Na Figura 3, é apresentado um modelo esquemático de integração entre soluções de base de dados que buscam atender às necessidades citadas.

Em iniciativas de desenvolvimento de bases de dados para o planejamento de transportes, um método de trabalho para sua construção é fundamental. Conforme apresentado por Date (2003), sem um método, a evolução e a manutenção da base de dados poderão ser executadas de forma errônea. A utilização de um método faz com que um conjunto mínimo de atividades sejam executadas e, dessa forma, é

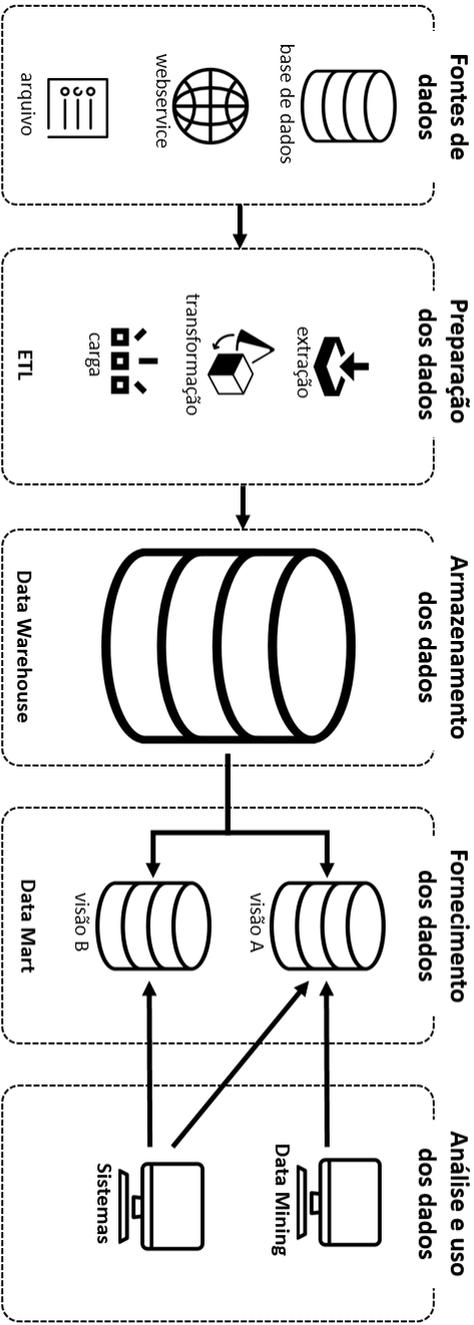


Figura 3 – Interação entre soluções de base de dados.  
 Fonte: Adaptada de DAMA (2012).

possível atingir os objetivos de uso traçados para a base de dados a ser construída.

### 2.1.1 Processo ETL

A sigla ETL vem do inglês *Extract, Transform and Load* (Extrair, Transformar e Carregar) e representa um processo cujas etapas são voltadas à coleta de dados de diversos sistemas (*data source*) e ao armazenamento em uma base de dados principal (*data warehouse* ou *data marts*) (Figura 4).

O processo ETL possui vastas aplicabilidades, podendo ser utilizado em qualquer iniciativa que apresente necessidade por transformação e tratamento de dados. No entanto, mostra-se essencialmente fundamental quando a carga de dados a ser realizada é muito grande.

Isso ocorre porque, quanto executado de forma planejada e estruturada (identificando compatibilidade entre estruturas, tipos de dados e origem e destino das informações), o ETL permite que o processo de carga de dados seja realizado sempre de forma padronizada, além de auxiliar a contornar problemas que poderiam invalidar uma base de dados, como povoamento de registros de tabelas erradas, inversão de colunas que possuem o mesmo tipo de dados e remoção de registros válidos.

Quando associados, os conceitos de ETL, *data warehouse*, *data mart* e *data mining* tendem a gerar um base de dados expansível, atualizável, estruturada e com regras bem definidas, e com dados são passíveis de utilização por diversos sistemas e usuários. Portanto, neste trabalho, é fundamental a análise e a explanação acerca do processo ETL.

Segundo Casters, Bouman e Dongen (2010), as principais etapas do processo ETL podem ser descritas como:

1. **Extrair:** etapa na qual ocorre o processamento necessário para conectar a várias fontes de dados, extrair os dados dessas fontes e disponibilizá-los para as etapas de processamento subsequentes.
2. **Transformar:** etapa na qual são aplicadas funções aos dados, a qual ocorre entre a extração de fontes e o carregamento em repositório de dados alvo. Essas funções podem conter (mas não estão limitadas a) as seguintes operações:

- movimento de dados;

- validação de dados contrarregras de qualidade de dados;
  - modificação de conteúdo ou estrutura dos dados;
  - integração dos dados com dados de outras fontes; e
  - cálculo de valores derivados ou agregados com base em dados processados.
3. **Carregar:** etapa na qual são realizados os processamentos necessários para carregar os dados em um repositório de dados alvo.

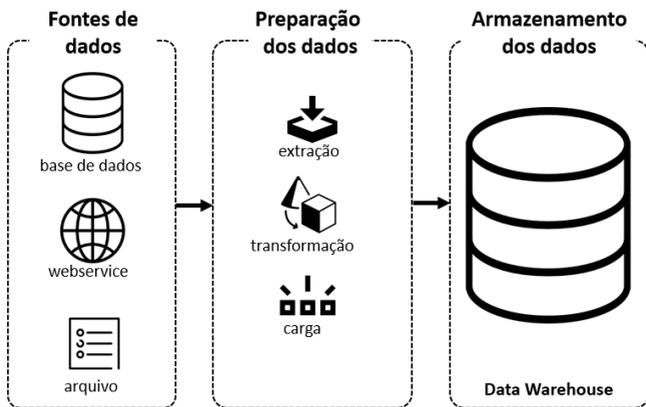


Figura 4 – A arquitetura convencional do ETL.

Fonte: Adaptada de Zhao e Huang (2006).

Segundo Kimball e Caserta (2004), as fontes de dados para uso do ETL não necessitam ter como origem, obrigatoriamente, sistemas propriamente ditos. Os dados podem ser originários de *sites*, arquivos de texto, banco de dados, *e-mail*, planilhas ou banco de dados. Casters, Bouman e Dongen (2010) complementam que a etapa de extração dos dados pode ser um empasse na modelagem e na concepção de um ETL devido às particularidades de cada formato em que os dados se encontram. Dupor e Jovanovic (2014) corroboram que a complexidade de uma implementação específica configura-se dependendo das fontes de dados, das regras de negócios, de *software* existente e de aplicativos de relatórios. O desafio é manter o foco na simples missão geral do sistema ETL.

As necessidades de integração de dados existirão enquanto os dados estiverem disponíveis em um formato digital. Nesse contexto,

pode-se utilizar um processo ETL, o qual pode ser concebido realizando codificação manual ou mesmo por meio de ferramentas computacionais específicas para esse fim.

Para Vaisman e Zimanyi (2014), quando de sua adoção de forma manual, é importante levar em consideração os custos associados ao desenvolvimento e à manutenção, pois os processos ETL são complexos e onerosos. Para garantir a redução do tempo e custo durante sua concepção, os autores sugerem a realização de uma modelagem no nível conceitual. No entanto, não existe um modelo conceitual definido para especificar tais processos; de qualquer modo, autores como Akkaoui et al. (2012), Oliveira e Belo (2013) e os próprios Vaisman e Zimanyi (2014) sugerem o uso da notação BMPN (*Business Process Model and Notation*) como alternativa para a modelagem.

Embora a construção manual do ETL seja onerosa, segundo Casters, Bouman e Dongen (2010), cerca de 45% de todo o trabalho ETL ainda é conduzido usando programas ou *scripts* codificados manualmente. Para os autores, as principais desvantagens da codificação manual são:

- propensa a erro;
- lenta em termos de tempo de desenvolvimento;
- manutenção complexa;
- falta de metadados; e
- falta de registros consistentes ou tratamentos de erros.

A opção da codificação manual era feita quando ferramentas ETL tinham valores bastante altos; mas atualmente, existem muitas alternativas de código aberto e outras de baixo custo disponíveis. Assim, depois de estabelecido o processo, não há nenhum ponto em tarefas de ETL que efetivamente demandem codificação manual.

Isso posto, existem variadas ferramentas computacionais que permitem a concepção de um processo ETL de maneira otimizada. Casters, Bouman e Dongen (2010) citam as ferramentas Microsoft Integration Services e Pentaho Data Integration (também conhecida como Kettle). Segundo os autores, são as mais recomendadas, pois possuem sua própria linguagem de programação para definir processos ETL. Para Vaisman e Zimanyi (2014), além das alternativas proprietárias, existem muitas de código aberto, não havendo nenhum ponto em tarefas de ETL em que a codificação manual seja vantajosa.

Segundo Casters, Bouman e Dongen (2010), o ETL não é usado somente para a carga de dados em repositórios de dados; pode servir para a seleção de informações com entrada para modelos de *data mining* ou para o carregamento de dados em sistemas de processamento em tempo real. Dupor e Jovanovic (2014) complementam que o ETL também é utilizado para tratar, limpar e transformar os dados ao longo do seu ciclo de vida.

Vaisman e Zimanyi (2014) destacam a importância de um projeto lógico na etapa de transformação, pois nesse ponto, todas as transformações dos dados de origem devem ser consideradas. Em alguns casos, os tratamentos podem ser diretos – por exemplo, a separação de endereços em seus componentes (rua, cidade, código postal) ou a extração de componentes de dado (mês e ano). Outras transformações podem exigir novas decisões – por exemplo, recalcular valores de medida para expressá-los em euros ou dólares ou para usar a moeda original e incluir a taxa de câmbio. Para os autores, é importante ficar claro que, em situações reais, podem ser necessárias transformações complexas de dados.

Uma vez que os dados que podem ser incluídos têm diferentes sistemas de origem, a questão das inconsistências deve ser algo a ser considerado no projeto, ou seja, uma estratégia apropriada para resolvê-las deve ser planejada. Portanto, os desenvolvedores de ETL devem projetar as estruturas de dados necessárias considerando todos os elementos para os quais deseja-se manter os tratamentos. Para a construção de ETL, é importante a definição de uma sequência preliminar de sua execução. Tal ação garante que todos os dados serão transformados e incluídos, com a consistência sendo garantida.

### **2.1.2 Data Warehouse**

A demanda por informações que permitam às organizações obterem vantagens estratégicas e operacionais é algo comum no meio corporativo. A partir dos anos 1990, a importância da análise de dados teve um aumento significativo, pois as organizações de todos os setores da indústria buscam tais informações como alternativa para manter sua vantagem competitiva (VAISMAN; ZIMANYI, 2014).

Segundo Malinowski e Zimanyi (2011), os requisitos necessários para a realização das análises de dados não são atendidos pelos bancos de dados tradicionais. Para os autores, a necessidade de garantir acesso rápido aos dados em um cenário de múltiplos usuários não é super-

tada pelos sistemas de bancos de dados tradicionais. O atendimento a essas demandas requer controle de concorrência, processamento de transações e aplicação de técnicas de recuperação, as quais garantem a consistência dos dados.

Os sistemas tradicionais de bancos de dados são conhecidos como bancos de dados operacionais ou sistemas de processamento de transações *on-line* (MALINOWSKI; ZIMANYI, 2011). Conforme Vaisman e Zimanyi (2014), normalmente bancos de dados operacionais contemplam dados detalhados e não históricos e, na maioria dos casos, possuem uma estrutura de dados altamente normalizada. Esse tipo de estrutura de dados, quando submetido a consultas complexas, envolvendo grande volume de informações, necessita juntar resultados de muitas tabelas ou ainda agregar grandes volumes de dados.

Em contraponto, Malinowski e Zimanyi (2011) citam que, existindo a necessidade de realização de análises de dados acerca do comportamento de uma organização como um todo por parte dos usuários, os dados operacionais devem ser integrados. Porém, a integração pode ser dificultosa, isso por causa das divergências na definição e no conteúdo de dados. Conforme descrito em Vaisman e Zimanyi (2014), os *data warehouses* foram propostos com a finalidade de melhor responder às crescentes demandas dos usuários por análises de dados para apoio à tomada de decisão.

Kimball's (1996), Date (2003), Kimball e Caserta (2004), Malinowski e Zimanyi (2011), Vaisman e Zimanyi (2014) e Shaw et al. (2016) afirmam que um *data warehouse* pode ser definido como um depósito de dados orientado por assunto, integrado, não volátil e variável em tempo de apoiar a tomada de decisões gerenciais. Conforme explicitado por Malinowski e Zimanyi (2011) e Vaisman e Zimanyi (2014), as características de um *data warehouse* são:

- **Orientado por assunto:** significa que os *data warehouses* se concentram em atender a necessidades analíticas de diferentes áreas de conhecimento de uma organização. Essas áreas variam dependendo dos tipos de atividades realizadas pela organização. Já em bases de dados operacionais, o foco está em funções específicas que as aplicações devem executar.
- **Integrado:** significa que os dados obtidos de vários sistemas operacionais e externos devem ser unidos, o que requer resolver problemas devido a diferenças na definição e no conteúdo de dados, tais como diferenças em formato e codificação de dados; sinônimos (campos com nomes diferentes, mas com dados iguais);

homônimos (campos com nomes iguais, mas com significados diferentes); e multiplicidade de ocorrências de dados. Nas bases de dados operacionais, esses problemas são normalmente resolvidos na fase de projeto.

- **Não volátil:** significa que a durabilidade dos dados é assegurada por não permitir a modificação e a remoção deles, ampliando o escopo dos dados para um período de tempo mais longo do que os sistemas operacionais. Um *data warehouse* reúne dados abrangendo vários anos (tipicamente de cinco a 10 anos ou além), enquanto que os dados em bases operacionais geralmente são mantidos apenas por um curto período de tempo (por exemplo, entre dois e seis meses), conforme necessário para operações diárias.
- **Variáveis em tempo:** indica a possibilidade de retenção de valores diferentes para a mesma informação, bem como o momento em que as mudanças dos valores ocorreram. Em contrapartida, um banco de dados operacional pode não ter suporte à temporalidade dos dados de forma explícita, uma vez que esta pode não ser necessária para as operações diárias. Além disso, a questão da temporalidade dificulta a implementação.

De acordo com Date (2003), comumente, as organizações apresentam a necessidade de soluções que tolerem a execução de demandas relacionadas à consulta de dados, isso sem impacto sobre seus sistemas operacionais. Associado a isso, o autor comenta que o volume de carga de trabalho dos *data warehouses* está diretamente associado ao apoio à decisão, o qual faz uso intensivo de consultas. Os bancos de dados operacionais tendem a ser grandes (frequentemente acima de 500 GB, crescendo cerca de 50% em um ano). Como resultado, o ajuste de desempenho de um banco de dados cujo volume de informações cresce pode ser um problema. Alinhado a isso, o autor cita alguns problemas que podem estar associados à questão de carga de trabalho:

- erros de projeto de bancos de dados;
- uso ineficiente de operações relacionais;
- fraqueza na implementação do modelo relacional do SGBD;
- falta de escalabilidade do próprio SGBD; e
- erros de projeto de arquitetura que limitam a capacidade e impedem a escalabilidade da plataforma.

Segundo Vaisman e Zimanyi (2014), a disponibilidade de enormes quantidades de dados exige uma mudança na forma como as práticas de armazenamento de dados e de inteligência de negócios são realizadas desde a década de 1990. Conforme Malinowski e Zimanyi (2011), em certos tipos de aplicativos de inteligência de negócios, os dados devem ser reestruturados para darem respostas em tempo hábil. Já na abordagem tradicional, em que os dados comerciais diários produzidos em uma organização são coletados em um enorme repositório comum para análise, os processos precisam ser revisados para darem conta de um tratamento eficiente para dados de larga escala. Em muitas organizações emergentes, nas quais as práticas de inteligência de negócio estão ganhando aceitação, fontes de dados de grande escala vêm tornando-se comuns, colocando novos desafios frente a comunidade de pesquisa de *data warehouse*.

Um *data warehouse* não deve ser visto como uma alternativa para a substituição de uma base de dados operacional como um todo, mas sim como uma estratégia para a evolução da otimização de extração de análises em dados estratégicos. Comumente, bases de dados operacionais são utilizadas como fonte de dados para a alimentação de *data warehouses*, sendo o processo de cargas das informações realizado em períodos de tempo determinados. Holanda et al. (2006) explicam que, devido ao fato de o *data warehouse* não ser volátil, quando os dados são inseridos, são mantidos por um período extenso, possibilitando acesso a dados históricos.

Para Malinowski e Zimanyi (2011), os dados, quando retirados de diversas fontes, devem ser transformados para o modelo do *data warehouse* e, em seguida, carregados no *data warehouse*. Conforme descrito pelos autores, esse processo é chamado de extração, transformação e carregamento (ETL), e foi provado crucial para o sucesso de um projeto de *data warehousing*.

Comumente, um *data warehouse* destina-se a analisar os dados de uma organização inteira. Muitas vezes, as áreas de uma organização requerem apenas uma parte do *data warehouse* organizacional especializado para suas necessidades. Esses armazéns de dados departamentais são chamados de *data marts*. Tais dados não são necessariamente restritos a um departamento específico, podendo existir compartilhamento com outras partes interessadas da organização.

Segundo Teorey et al. (2013), um *data warehouse* tem um ciclo de vida (Figura 5) que se inicia com um diálogo para o estabelecimento do plano de projeto e para a identificação das necessidades de negócio de tal plano. Para a construção de um *data warehouse*, deve haver

um planejamento, tendo-se, como resultado, as necessidades de negócio para o *data warehouse*. Em seguida, é elaborado um detalhamento do plano de implementação, o qual se subdivide em três atividades em paralelo:

1. **Projeto arquitetônico:** nesta atividade, as questões de plataforma são tratadas, incluindo o projeto técnico arquitetônico e a seleção e a instalação de produto.
2. **Modelagem dimensional:** nesta atividade, as questões de dados são abordadas, incluindo a modelagem dimensional e, depois, o projeto físico, seguido por projeto e desenvolvimento da preparação de dados.
3. **Especificação da aplicação analítica:** nesta atividade, as necessidades analíticas especiais dos usuários são atendidas, incluindo a especificação da aplicação analítica, seguida pelo desenvolvimento da aplicação analítica.

Após a conclusão das três atividades de projeto, inicia-se a implementação, que é o momento em que a construção do *data warehouse* propriamente dita acontece. Por fim, tem-se a manutenção e a detecção de pontos de mudança nos requisitos do *data warehouse*. Caso necessário, ajustes são planejados por meio da repetição do ciclo de vida. Porém, se o sistema tornar-se inativo, o ciclo de vida termina.

Date (2003), Malinowski e Zimanyi (2011) e Vaisman e Zimanyi (2014) concordam que um *data warehouse* pode ser visto como uma coleção de *data marts*. Isso representa uma visão *bottom-up* em sua concepção, na qual um *data warehouse* é construído a partir de *data marts*, os quais são integrados para que se obtenha o *data warehouse*. Essa abordagem tem a finalidade de atender organizações que não estão dispostas a correrem riscos durante a construção de um *data warehouse* – que pode demorar muito para ser concluído – ou que necessitam de resultados rápidos. Por outro lado, na visão clássica do *data warehouse*, os *data marts* são obtidos a partir do *data warehouse* de uma forma *top-down*. Nessa abordagem, um *data mart* é apenas uma visão lógica de um *data warehouse*.

Segundo Date (2003), a análise de dados para a exploração dos conteúdos de um *data warehouse* com a finalidade de fornecer informações essenciais ao processo de tomada de decisão pode ser realizada por meio de *data mining*, que consiste em técnicas estatísticas para análise dos dados de um *data warehouse*, a fim de descobrir conhecimentos úteis e que não são fáceis de obter-se a partir dos dados originais.

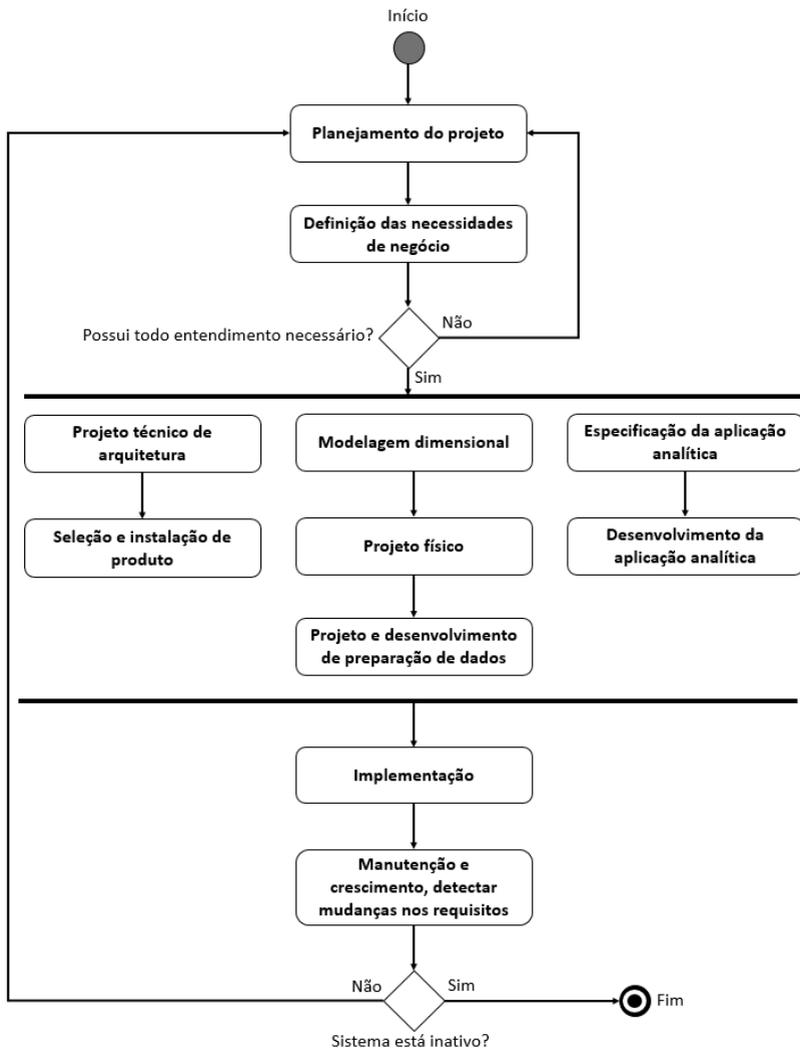


Figura 5 – Ciclo de vida do *data warehouse*.  
 Fonte: Teorey et al. (2013).

Desenvolver um *data warehouse* é uma iniciativa complexa e que precisa ser cuidadosamente planejada, principalmente quanto à utilização de informações de bancos de dados operacionais existentes. Além disso, durante a fase de projeto, é necessário levar em consi-

deração as etapas de especificação de requisitos, *design* conceitual, *design* lógico e *design* físico. Em cada etapa, é necessário analisar os requisitos dos usuários, verificando sistemas de origem ou combinando ambos. Portanto, seu desenvolvimento exige um planejamento quanto às especificidades do negócio a ser tratado, observando cada contraponto aos benefícios de *data warehouse* para a organização.

### 2.1.3 Data Mart

Segundo Kimball (1996), pode-se definir *data mart* como um conjunto de dados baseados em uma base de dados atômica, que não depende de dados físicos e que tem a finalidade de extrair informações de uma base de dados operacional. Westerman (2000) define um *data mart* com um *data warehouse*, mas para um subconjunto menor de elementos de dados (Figura 6). Os autores Westerman (2000), Kimball e Caserta (2004), Inmon (2005), (ARFAOUI; AKAICHI, 2013) e Prema, Clara e Pethalakshmi (2014) concordam que, de modo geral, um *data mart* é uma base de dados destinada ao uso e ao atendimento das necessidades do usuário final.

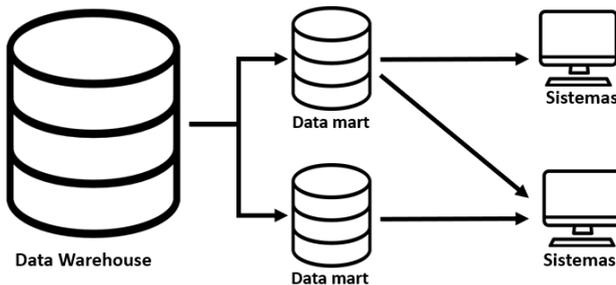


Figura 6 – *Data mart*.

Fonte: Adaptada de Westerman (2000).

Conforme Prema, Clara e Pethalakshmi (2014), um *data mart* pode armazenar informações que abordem as necessidades táticas e estratégicas de uma organização, permitindo que as suas principais funções operacionais tenham um desempenho adequado quanto aos processos de negócio a serem gerenciados. Ainda segundo os autores, existem cinco motivos para construir-se um *data mart*, sendo eles:

1. melhora o tempo de resposta do usuário final;
2. cria uma visão coletiva para um grupo de usuários;
3. tem criação facilitada;
4. fácil acesso a dados frequentemente necessários; e
5. é menos custoso que implementar um *data warehouse* completo.

Segundo Kimball e Caserta (2004), um projeto de construção de *data warehouse* é bem-sucedido, quando cada *data mart* é considerado um projeto separado, com seu próprio orçamento e cronograma. Assim, o *data warehouse* de uma organização evolui e cresce com a conclusão de cada projeto de *data mart*. Para os autores, nesse modelo, uma boa alternativa é pensar o *data warehouse* como um processo, não como um projeto.

Kimball e Ross (2002) contribuem sugerindo que os *data marts* armazenem dados de forma resumida; assim, irão permitir a navegação facilitada sobre os dados. De acordo com os autores, o resultado dessa iniciativa é a diminuição de tempo de acesso aos dados devido à menor quantidade de tabelas e relacionamentos. Kimball e Caserta (2004) comentam que *data marts* construídos de acordo com as definições equivocadas não serão capazes de lidar com a mudança de requisitos de negócios, pois os detalhes foram resumidos ou não considerados no projeto.

Conforme descrito por Date (2003), Malinowski e Zimanyi (2011) e Vaisman e Zimanyi (2014), um conjunto de *data marts* permite a construção de um *data warehouse* e vice-versa; portanto, um *data mart* pode ser construído a partir da especificação de um *data warehouse*. Westerman (2000) afirma que a maioria das empresas inicia seus projetos de *data warehouse* construindo primeiro um *data mart*. Entretanto, Inmon (2005) comenta que, nessa abordagem, muitos problemas se tornam evidentes, tais como:

- o número de programas de extração cresce;
- cada banco de dados novo deve ser compatibilizado com a base de dados legado;
- não há estrutura compatível com ajustes futuros; e
- grande volume de dados redundantes.

Segundo Westerman (2000), normalmente os *data marts* compõem conjuntos menores de dados ou uma única área de assunto, sendo projetados para um grupo especial de usuários. Os *data marts* podem ou não estar relacionados diretamente com outros *data marts* ou outro *data warehouse*; dessa forma, os *data marts* podem atuar de forma independente. Tal característica não implica que *data marts* sejam mais baratos ou fáceis de construir-se; seu desenvolvimento pode ser tão complexo e demorado quanto o de um *data warehouse*. Para Westerman (2000), uma empresa, ao construir muitos *data marts* – um para cada subconjunto de dados –, teria um aumento de complexidade. Porém, às vezes, essa iniciativa pode melhorar o desempenho de consulta e tempo de resposta quanto à requisição de dados. Um *data mart* deve ser construído para resolver problemas de negócio e não técnicos; assim, quando são elaborados para gerenciar um grande volume de dados, o negócio sofre com a manutenção dessas informações.

#### 2.1.4 *Data Mining*

Com o grande volume de informações que as organizações vêm armazenando ao longo do tempo, surgem a necessidade e a dificuldade de extração de informações úteis de uma grande base de dados. Segundo Wojciechowski (2015), muitas das informações importantes para os negócios das organizações encontram-se escondidas ou espalhadas nas bases de dados existentes, e a adoção de mecanismos de consultas simples não consegue extrair informações com utilidades reais.

Segundo Vaisman e Zimanyi (2014), a técnica de *data mining* surgiu com o aumento da capacidade de coletar e processar dados e foi aprimorada com as possibilidades oferecidas pelo armazenamento de dados. Assim, com base no vasto volume de informações, surgiu a necessidade de extrair informações úteis. Elmasri e Navathe (2005), Malinowski e Zimanyi (2011) e Vaisman e Zimanyi (2014) corroboram que *data mining* é uma técnica na área de banco de dados que busca a análise de um conjunto de dados, frequentemente grande, com a finalidade de encontrar relacionamentos, padrões, tendências, resumindo os dados de forma que agreguem valor para uma organização.

Kimball e Ross (2002) complementam que *data mining* é um conjunto de consultas não direcionadas, muitas vezes voltadas a dados atômicos, e que buscam encontrar padrões inesperados nos dados. Para os autores, geralmente, a técnica de *data mining* está inserida no contexto do *data warehouse*, isso por meio de um cliente deste último.

Elmasri e Navathe (2005) complementam que, para as duas técnicas serem utilizadas paralelamente de forma eficiente, o *data warehouse* deve ter um conjunto de dados agregados ou sumarizados.

*Data mining* é um dos passos necessários para a identificação de conhecimento em banco de dados, visando à extração de informações não triviais, implícitas, anteriormente desconhecidas e potencialmente úteis de dados em bancos de dados. Para Vaisman e Zimanyi (2014), o processo de identificação de conhecimento envolve várias fases, tais como limpeza de dados, seleção, transformação, redução, seleção de modelos e, finalmente, exploração do conhecimento extraído. Chapman et al. (2000) propõem um processo cíclico (Figura 7) para a construção de um *data mining*, no qual as etapas são flexíveis, podendo existir a necessidade do retorno entre as etapas propostas. As seis etapas do processo são:

1. **Entendimento do negócio:** nesta etapa, é realizada uma atividade para entendimento de negócios, objetivos, requisitos e limitações acerca das aplicações de um *data mining*.
2. **Entendimento dos dados:** a partir da base de dados existente, os dados são selecionados e suas características, restrições quantitativas e qualitativas são utilizadas para o estabelecimento de definições sobre os subconjuntos de dados. Quando necessário, é possível retornar à etapa anterior, objetivando um melhor entendimento sobre o negócio.
3. **Preparação dos dados:** são realizadas a seleção, a preparação, a limpeza e a tradução dos dados, a fim de realizar a carga dos dados no *data mining*.
4. **Modelagem:** são aplicadas técnicas sobre os dados com a finalidade de modelar o *data mining* para a manipulação dos dados. Por vezes, quando forem encontrados novos dados, é sugerido retorno à fase anterior.
5. **Avaliação:** nesse momento, o modelo estabelecido deverá ser avaliado no que diz respeito à sua aplicação para o problema sugerido, o que pode provocar revisões nos objetivos de negócios.
6. **Implantação:** esta etapa consiste na aplicação do modelo desenvolvido na produção dos dados desejados, avaliando os resultados de forma aderente aos objetivos estabelecidos.

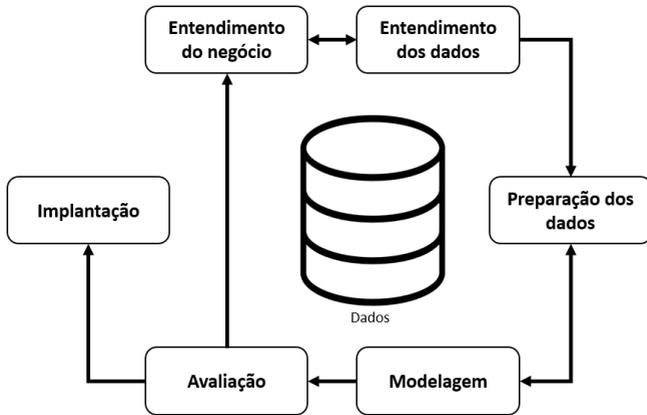


Figura 7 – Etapas do modelo de referência CRISP-DM.  
 Fonte: Adaptada de Chapman et al. (2000).

Os resultados do *data mining* podem ser mostrados em uma variedade de formatos, como listagens, gráficos, tabelas resumidas ou gráficos. Para Elmasri e Navathe (2005), a identificação de conhecimento permite verificar a evolução de um dado para informação e de informação para conhecimento, isso ao passo que evolui o processamento das informações. Para os autores, é comum documentar o conhecimento identificado durante o *data mining* da seguinte forma:

1. **Regras de associação:** relacionam a presença de um conjunto de itens com a faixa de valores de outro conjunto de variáveis.
2. **Hierarquias de classificação:** o objetivo é trabalhar em um conjunto de eventos ou transações para criar uma hierarquia de classes.
3. **Padrões sequenciais:** uma sequência de ações ou eventos é investigada.
4. **Padrões com séries temporais:** similaridades podem ser encontradas em posições de uma série temporal de dados, que é uma sequência de dados formada por intervalos regulares, como vendas diárias ou preço diário de fechamento de ações.
5. **Agrupamento de dados:** dada uma população de eventos particionados (segmentados) em conjuntos de elementos similares.

Segundo Teorey et al. (2013), comumente são utilizadas duas técnicas para a identificação de conhecimento em um baco de dados por meio de *data mining*. Na primeira técnica, o usuário aplica análise estatística via consultas em um banco de dados padrão, tentando confirmar ou refutar hipóteses impostas. Na segunda técnica, são realizadas correlações entre os dados para que eles apresentem hipóteses promissoras; assim, o usuário leva-as em consideração durante a análise. Para Elmasri e Navathe (2005), quando o *data mining* é utilizado a fim de confirmar ou refutar hipóteses ao invés de permitir a identificação das hipóteses, pode não apresentar resultados consistentes para os seguintes aspectos:

1. **Predição:** o *data mining* pode não mostrar como certos atributos dos dados irão comportar-se no futuro.
2. **Identificação:** os padrões de dados podem ser usados para identificar a existência de um item, um evento ou uma atividade.
3. **Classificação:** o *data mining* pode particionar os dados, podendo as diferentes classes ou categorias ser identificadas baseadas em combinações de parâmetros.
4. **Otimização:** um objetivo do *data mining* pode ser voltado à otimização do uso de recursos limitados, tais como tempo, espaço, dinheiro ou materiais, além de maximizar as variáveis de saída.

No processo de construção de um *data mining*, devem ser considerados aspectos relacionados tanto às fontes de dados quanto aos usuários de interesse. No processo de construção é que se define o algoritmo, levando em consideração os custos e sua viabilidade de aplicação. Na prática, os algoritmos de *data mining* são aplicados para tentar solucionar diversos problemas do dia a dia das organizações.

Para Barbieri (2011), no processo de construção do *data mining*, as técnicas aplicadas são baseadas em fundamentos estatísticos. Wojciechowski (2015) aponta que o *data mining* está contextualizado em um processo conhecido como KDD (*Knowledge Discovery in Database*), reconhecendo padrões em repositórios de dados. Barbieri (2011) destaca que as principais técnicas são Árvore de decisão, Análise de conglomerados, Redes neurais (não é exatamente uma técnica estatística, mas um recurso matemático/computacional que pode ser usado no contexto do *data mining*), Análise de regressão (linear e não linear) e Séries temporais.

Por ser um processo de exploração de dados em um grande volume de informações, o *data mining* está presente em vários setores e áreas do conhecimento, como detecção de fraudes, identificação de estratégias lucrativas de vendas, avaliação e eficácia de promoções. Segundo Wojciechowski (2015), diariamente, as organizações analisam seus bancos de dados em busca de informações de negócio relevantes, as quais, embora sempre estivessem lá, não eram conhecidas.

## 2.2 DADOS ABERTOS GOVERNAMENTAIS

Dados abertos governamentais são informações que estão disponíveis de forma livre para a sociedade, não apenas para leitura e acompanhamento, mas sim para reutilização em projetos e estudos de quaisquer naturezas. Tal tema teve início por volta de 2009, principalmente em países como Estados Unidos, Reino Unido, Canadá e Nova Zelândia. Atualmente, muitos desses dados que estão abertos são coletados por pessoas e empresas a fim de que possam executar suas atividades de estudos e análises. Porém, trata-se de um recurso ainda pouco explorado pela sociedade (W3C, 2017).

O governo brasileiro vem discutindo, há muitos anos, alternativas para a viabilização da transparência dos dados. Em 2011, o governo passou a integrar a Parceria para Governo Aberto (Open Government Partnership) e, no mesmo ano, sancionou a Lei de Acesso à Informação (Lei n. 12.527/2011), a qual estabelece que todos os órgãos públicos têm a obrigatoriedade de divulgar dados de interesse coletivo (TCU, 2015). Segundo W3C (2017), existem cinco motivos para que os dados estejam abertos:

- transparência na gestão pública;
- contribuição à sociedade com serviços inovadores;
- melhoria da qualidade dos dados governamentais;
- viabilização de novos negócios; e
- obrigatoriedade por lei.

Para viabilizar a Lei de Acesso à Informação, o governo criou diversos *sites* e portais para que as entidades disponibilizem seus dados, sendo o principal deles o Portal Brasileiro de Dados Abertos (TCU, 2015). A disponibilização de dados abertos também é uma forma de a

administração pública prestar contas para a sociedade e muitos órgãos disponibilizam dados sem requerimento prévio do governo federal. Os dados somente são considerados abertos quando atendem aos seguintes requisitos (W3C, 2017):

- **São completos:** os dados precisam estar disponíveis de forma completa, sem necessidade de outros meios para a sua utilização e compreensão.
- **São primários:** os dados precisam estar da maneira como foram gerados/coletados, não tendo passado por processos de tratamento e/ou agregação.
- **Estão atualizados:** os dados precisam ser mantidos na versão mais atual possível, a fim de preservar sua utilidade.
- **São acessíveis:** os dados precisam estar disponíveis na internet de forma fácil e sem restrições.
- **São processáveis por máquina:** os dados precisam estar padronizados de maneira que se possa utilizar rotinas automatizadas para seu processamento.
- **Não é necessária a identificação para acessá-los:** os dados precisam estar disponíveis sem restrição de acesso ou necessidade de identificação do interessado.
- **São disponibilizados em formatos não proprietários:** os dados precisam estar disponíveis em formatos sem restrição de licença e sem necessidade de *software* proprietário para seu uso.
- **São livres de licenças:** os dados precisam não devem estar sujeitos a restrição de uso, independentemente do seu tipo.

A política de liberação dos dados governamentais funciona como um instrumento de transformação social, permitindo que a população avalie e fiscalize as ações do governo. Na academia, tais dados viabilizam a execução de estudos e projetos que auxiliam na evolução e no crescimento do país. Nesse contexto, no Portal Brasileiro de Dados Abertos, estão disponíveis para *download* em formato aberto diversos conjuntos de dados relevantes para o interesse público (TCU, 2015).

### 2.2.1 Instituto Brasileiro de Geografia e Estatística

O Instituto Brasileiro de Geografia e Estatística (IBGE) é o mais importante instrumento brasileiro para obtenção de dados e informações, visando atender a diversos segmentos da sociedade. Entre suas atribuições, estão a produção e a análise de informações estatísticas, a coordenação e a consolidação das informações estatísticas, a produção e a análise de informações geográficas, a coordenação e a consolidação das informações geográficas, a estruturação e a implantação de um sistema de informações ambientais, documentação e disseminação de informações, além da coordenação dos sistemas estatístico e cartográfico nacionais (IBGE, 2017b).

O IBGE possui uma ferramenta chamada Cidades, cujo objetivo é disponibilizar informações acerca de todos os municípios brasileiros de forma centralizada. Nessa ferramenta, estão dispostas tabelas, gráficos, históricos e mapas que traçam um perfil sobre os municípios (IBGE, 2017a). É disponibilizada, inclusive, uma planilha para *download* que contém as seguintes informações sobre os municípios:

- código da unidade da federação;
- nome da unidade da federação;
- código do município;
- nome do município;
- nome da região metropolitana;
- código da mesorregião;
- nome da mesorregião;
- código da microrregião;
- nome da microrregião;
- ano de referência (para os indicadores);
- valor adicionado bruto da agropecuária, a preços correntes;
- valor adicionado bruto da indústria, a preços correntes;
- valor adicionado bruto dos Serviços, a preços correntes – exclusive administração, saúde e educação públicas e seguridade social;

- valor adicionado bruto da administração, saúde e educação públicas e seguridade social, a preços correntes;
- valor adicionado bruto total, a preços correntes;
- impostos, líquidos de subsídios, sobre produtos, a preços correntes;
- Produto Interno Bruto, a preços correntes;
- população (quantidade de habitantes); e
- Produto Interno Bruto *per capita*.

Ademais, o IBGE realiza de forma periódica um levantamento detalhado sobre as informações dos municípios brasileiros. O resultado é disponibilizado à sociedade por meio de uma Pesquisa de Informações Básicas Municipais (MUNIC), a qual engloba informações dos 5.570 municípios brasileiros desde de 1999, oferecendo aos cidadãos dados estatísticos e cadastrais que formam um conjunto relevante de indicadores de avaliação e monitoramento do quadro institucional e administrativo referente aos municípios brasileiros (IBGE, 2017b).

### **2.2.2 Agência Nacional de Aviação Civil**

A Agência Nacional de Aviação Civil (ANAC) é uma agência reguladora federal que tem o objetivo de regular e fiscalizar as atividades de aviação civil e a infraestrutura aeroportuária do país. Atua para prover segurança na aviação civil, estimular a concorrência e a melhoria na prestação de serviços do setor. Desse modo, o trabalho da ANAC consiste em elaborar normas, certificar empresas, oficinas, escolas, profissionais da aviação civil, aeródromos e aeroportos e fiscalizar as operações de aeronaves, de empresas aéreas, de aeroportos e de profissionais do setor e de aeroportos, com foco na segurança e na qualidade do transporte aéreo (ANAC, 2017f).

Como resultado das suas atividades de regulação e fiscalização, a ANAC libera, de forma pública, dados, os quais estão disponíveis no portal da Agência (<http://www.anac.gov.br>) para o acesso de qualquer cidadão. Entre eles, pode-se citar dados de cadastro de aeródromos, dados estatísticos, dados de horário de transporte aéreo e registros de voos regulares.

### 2.2.2.1 Cadastro de aeródromos

Segundo a Lei n. 7.565, de 19 de dezembro de 1986, aeródromo é toda área destinada a pouso e decolagens de aeronave. O sistema aeroportuário é composto pelo conjunto de aeródromos brasileiros, com todas as pistas de pouso, pistas de táxi, pátios de estacionamento de aeronave, terminais de carga aérea, terminais de passageiros e respectivas facilidades.

Os aeródromos são classificados em militares e civis, sendo os militares de uso exclusivo das Forças Aéreas Brasileiras e os civis normalmente destinados a usos civis e comerciais. Aeródromos civis poderão ser utilizados por aeronaves militares e aeródromos militares por aeronaves civis quando obedecidas as prescrições estabelecidas pela autoridade aeronáutica. Os aeródromos civis podem ainda ser classificados em privados e públicos (BRASIL, 1986).

Os aeródromos privados são de uso exclusivo do proprietário, portanto, este tem direitos legais para a execução e a aplicação de tarifas. A liberação de operação é concedida por meio de registro, podendo ser encerrada a qualquer tempo pelo proprietário e/ou pela ANAC (ANAC, 2017a).

Os aeródromos públicos são constituídos de uma área mantida e de posse da União. Sua operação é regida pela ANAC e requer um processo de homologação para a liberação de suas atividades; porém, seu fechamento apenas pode ser executado mediante processo administrativo. As áreas onde estão situados tais aeródromos estão sujeitas a restrições especiais em relação ao plano básico de zona de proteção de aeródromos e ao plano de zoneamento de ruídos. No conjunto de aeródromos públicos, existem alguns que são classificados como aeroportos. Para isso, devem possuir instalações e facilidades que dão suporte à operação aeroportuária (ANAC, 2017a).

”São facilidades: o balizamento diurno e noturno; a iluminação do pátio; serviço contra-incêndio especializado e o serviço de remoção de emergência médica; área de pré-embarque, climatização, ônibus, ponte de embarque, sistema de esteiras para despacho de bagagem, carrinhos para passageiros, pontes de desembarque, sistema de ascenso-descenso de passageiros por escadas rolantes, orientação por circuito fechado de televisão, sistema semi-automático anunciador de mensagem, sistema de som, sistema in-

formativo de voo, climatização geral, locais destinados a serviços públicos, locais destinados a apoio comercial, serviço médico, serviço de salvamento aquático especializado e outras, cuja implantação seja autorizada ou determinada pela autoridade aeronáutica” (BRASIL, 1986).

A ANAC disponibiliza em seu portal (<http://www.anac.gov.br>) duas listagens referentes a aeródromos: uma contendo aeródromos públicos e outra os privados. Nelas, é possível encontrar as seguintes informações (ANAC, 2017a):

- **Código ICAO:** código único, composto por quatro letras, que permite a identificação dos aeródromos. Tal informação é controlada pela Organização da Aviação Civil Internacional.
- **Tipo de aeródromo:** classifica se a informação é referente a um aeródromo ou a heliponto (no contexto deste trabalho, todos os dados são classificados como aeródromo);
- **Nome do aeródromo:** nome de registro do aeródromo.
- **Município:** para a listagem de aeródromos públicos, este campo identifica o município atendido pelo aeródromo. Na listagem de aeródromos privados, identifica o município onde o aeródromo está localizado.
- **Unidade Federativa:** dispõe a unidade da federação onde o aeródromo está localizado por meio da sigla da unidade.
- **Localização:** são disponibilizadas as coordenadas geográficas (latitude e longitude) no formato geodésico, além da altitude em metros.
- **Operação:** é disponibilizado o tipo de operação do aeródromo. Indica se o aeródromo opera com VFR (*Visual Flight Rules*) diurno e/ou noturno, e se opera com IFR (*Instrument Flight Rules*) diurno e/ou noturno.
- **Pistas:** sobre as pistas de pouso e decolagem, são disponibilizadas informações como designação (orientação das cabeceiras A e B), comprimento, largura, resistência (via nomenclatura de PCN - *Pavement Classification Number*, que indica tipo de pavimento, resistência, pressão admissível dos pneus da aeronave e método de avaliação realizado na pista) e tipo de superfície da pista (asfalto, concreto, grama ou terra).

Os dados de aeródromos públicos e privados estão disponíveis em listagens separadas com formatos (quantidade de colunas) diferentes. A planilha de aeródromos privados possui dados de somente uma pista, enquanto que na de aeródromos privados existem dados (quando disponíveis) de até três pistas.

#### 2.2.2.2 Horário de transporte

O horário de transporte (HOTRAN) é um documento emitido e aprovado pela ANAC que formaliza a autorização para a exploração de uma ou mais rotas aéreas regulares domésticas e internacionais de passageiros, cargas, rede postal e táxi aéreo. Sendo assim, tal documento apresenta informações detalhadas a respeito da linha autorizada para operação (ANAC, 2017d).

Todos o processo de autorização das linhas é disponibilizado, desde a solicitação de autorização até a autorização propriamente dita. Quando uma solicitação é feita (diretamente pela companhia aérea), ela é inserida na planilha de voos a vigorar (a ANAC identifica tal planilha como 1.4); após a aprovação (realizada pela ANAC) e a chegada da data de vigência, tal informação é transferida para a planilha de voos vigentes (a ANAC identifica tal planilha como 1.5) (ANAC, 2017d).

A informação de HOTRAN é de suma importância para a aviação civil brasileira, pois nesse conjunto de documentos está contido todo o planejamento de rotas, horários e frequência dos voos. Além disso, com essa informação, é possível determinar em quais períodos os aeroportos terão movimentação.

No conjunto de informações de voos vigentes, estão disponíveis os seguintes dados (ANAC, 2017c):

- **Código da empresa:** sigla única que identifica a empresa autorizada a realizar o voo.
- **Empresa:** nome da empresa aérea que está autorizada a realizar a operação.
- **Número do voo:** código numérico único que identifica o voo.
- **Equipamento:** código identificador do modelo de aeronave utilizado na operação.
- **Quantidade de assentos:** quantidade de assentos disponíveis no voo, informação que está diretamente relacionada ao modelo de aeronave a ser utilizado.

- **Número HOTRAN:** código que permite identificar a autorização de rota. Para um número de HOTRAN, podem existir uma ou mais rotas. O número do HOTRAN segue o seguinte formato:

CCC-000NNN-VVV, onde:

CCC: identifica o código identificador da companhia aérea;

NNN: identifica o número de HOTRAN; e

VVV: identifica a versão do HOTRAN.

- **Data solicitação:** data em que o HOTRAN foi solicitado.
- **Data aprovação:** data em que o HOTRAN foi aprovado.
- **Data Vigência:** data em que o HOTRAN passou a ter validade.
- **Natureza Operação:** classificação da natureza de operação do voo, podendo ser internacional, doméstico (representado por regional, especial e nacional), cargueiro doméstico, cargueiro internacional, rede postal ou táxi aéreo.
- **Número da etapa:** número que permite identificar a sequência de etapas de uma rota.
- **Código de origem:** código identificador do aeroporto de origem de uma etapa.
- **Aeroporto de origem:** nome do aeroporto de origem da etapa do voo.
- **Código de destino:** código identificador do aeroporto de destino de uma etapa.
- **Aeroporto de destino:** nome do aeroporto de destino da etapa do voo.
- **Horário de partida:** horário de partida do aeroporto de origem da etapa (decolagem).
- **Horário de chegada:** horário de chegada ao aeroporto de destino da etapa (pouso).
- **CODESHARE:** código criado para identificar o compartilhamento de rotas entre companhias aéreas.

- **Observação:** informações adicionais que a empresa aérea deseja informar.

O HOTRAN (voos vigentes) está disponível diariamente desde 2003. Tais informações podem ser acessadas diretamente do portal da ANAC (<http://www.anac.gov.br>) nos formatos HTML, XLS ou CSV (ANAC, 2017c).

### 2.2.2.3 Movimentação aeroportuária

Entre as competências da ANAC, estão a elaboração e a divulgação de estudos sobre o setor de transporte aéreo brasileiro. As informações e os dados coletados sobre o setor revelam-se de suma importância, pois permitem o bom desempenho das atividades, do planejamento e do desenvolvimento de tais estudos (ANAC, 2017b).

Segundo a ANAC (2017f), os dados são empregados no auxílio ao processo legislativo, bem como na elaboração de políticas públicas e no processo regulatório que abrange o setor, no planejamento de investimentos, na tomada de decisões estratégicas, na prospecção de mercado, no controle de concorrência, no planejamento de frota e no planejamento de infraestrutura aeroportuária e de navegação aérea. Com a finalidade de ampliar o conhecimento da sociedade, a ANAC tem divulgado tais informações sobre o setor em seu portal. No conjunto de dados, é possível encontrar as seguintes informações (ANAC, 2017b):

- **Empresa aérea:** empresa aérea responsável por operar as etapas.
- **Natureza do voo:** refere-se à natureza das etapas, podendo ser Doméstico (caso as etapas tenham o pouso e a decolagem realizados no Brasil e sejam operadas por empresas brasileiras) ou Internacional (caso contrário).
- **Aeródromo de origem:** refere-se ao aeródromo de início do trecho, ou seja, onde acontece a decolagem.
- **Aeródromo de destino:** refere-se ao aeródromo de fim do trecho, ou seja, onde acontece o pouso.
- **Tipo de voo:** refere-se ao tipo de operação das etapas, podendo ser Improdutivas (etapas que não geraram receita à empresa aérea), Regulares (etapas remuneradas que são realizadas sob uma numeração de Horário de Transporte) e Não regulares

(etapas remuneradas que não são realizadas sob uma numeração de Horário de Transporte).

- **Passageiros pagos:** refere-se aos passageiros que ocupam assentos comercializados ao público e que geram receita para a empresa de transporte aéreo.
- **Passageiros grátis:** refere-se aos passageiros que ocupam assentos comercializados ao público, mas que não geram receita para a empresa de transporte aéreo.
- **Carga paga:** refere-se à quantidade total, expressa em quilogramas, de todos os bens que tenham sido transportados na aeronave, exceto correio e bagagem, e que tenha gerado receita direta ou indireta para a empresa aérea.
- **Carga grátis:** refere-se à quantidade total, expressa em quilogramas, de todos os bens que tenham sido transportados na aeronave, exceto correio e bagagem, e que não tenha gerado receita direta ou indireta para a empresa aérea.
- **Correio:** refere-se à quantidade de objetos transportados de rede postal em cada trecho de voo realizado, expressa em quilogramas.
- **ASK (*Available Seat Kilometers*):** refere-se ao volume de assentos oferecidos por quilômetros, ou seja, a soma entre o número de assentos oferecido e a distância das etapas.
- **RPK (*Revenue Passenger Kilometres*):** refere-se ao volume de passageiros transportados por quilômetros, ou seja, a soma entre o número de passageiros pagos e a distância das etapas.
- **ATK (*Available Tonne Kilometres*):** refere-se ao volume de tonelada por quilômetros oferecidos, ou seja, a soma entre *payload*, que é a capacidade total de peso disponível na aeronave, expressa em quilogramas, para efetuar o transporte de passageiros, carga e correio; e a distância das etapas, dividido por 1.000.
- **RTK (*Revenue Tonne Kilometres*):** refere-se ao volume de toneladas por quilômetros transportados, ou seja, a soma entre os quilogramas carregados pagos, sendo que cada passageiro possui o peso estimado de 75 Kg, e a distância das etapas, dividido por 1.000.

- **Combustível:** refere-se à quantidade, em litros, de combustível consumida pela aeronave na execução da referida etapa. Informação disponível apenas para empresas brasileiras.
- **Distância:** refere-se à distância, expressa em quilômetros, entre os aeródromos de origem e destino da etapa, considerando a curvatura do planeta Terra.
- **Horas voadas:** refere-se ao número de horas de voo entre os aeródromos de origem e de destino da etapa.
- **Decolagens:** refere-se ao número de decolagens que ocorreram entre os aeródromos de origem e de destino da etapa.
- **Carga paga (km):** refere-se ao volume de carga paga (kg) em cada quilômetro, ou seja, a soma entre a quantidade (kg) de carga paga e a distância das etapas.
- **Carga grátis (km):** refere-se ao volume de carga grátis (kg) em cada quilômetro, ou seja, a soma entre a quantidade (kg) de carga grátis e a distância das etapas.
- **Correio (km):** refere-se ao volume de correio (kg) em cada quilômetro, ou seja, a soma entre a quantidade (kg) de correio e a distância das etapas.
- **Assentos:** é o número disponível de assentos em cada etapa de voo de acordo com a configuração da aeronave na execução da etapa.
- **Payload (kg):** é a capacidade total de peso na aeronave, expressa em quilogramas, disponível para efetuar o transporte de passageiros, carga e correio.

Os dados de movimentação aeroportuária estão disponíveis em formatos de planilhas; está disponível ainda uma série histórica a partir do ano de 2000, agrupada por rota, empresa, ano e mês. Os dados de ASK, RPK, ATK, RTK são apresentados consolidados por rota, considerando os valores registrados no mês. As atualizações ocorrem mensalmente e esses dados não possuem restrições de acesso, sendo abertos para a consulta de qualquer interessado (ANAC, 2017b).

#### 2.2.2.4 Voo regular ativo

Na base de dados de dados do voo regular ativo (VRA), são disponibilizados os registros da operação do transporte aéreo regular, apresentando, para cada rota, os horários de ocorrência e a situação. Além do registro de voos realizados, a base contempla os voos cancelados e atrasados; para eles, são disponibilizadas as justificativas da companhia aérea. A ANAC gera essa base de dados a partir do cruzamento da base de HOTRAN, que possui todos os voos regulares e não regulares vigentes com autorização para operação, com os boletins de alteração de voo (BAV), nos quais constam dados relativos a eventuais inclusões e alterações de voos (ANAC, 2017g).

Como citado, a operação aérea regular deve ser regida pelo HOTRAN. Nesse contexto, qualquer alteração na operação planejada deve ser indicada, como cancelamento de uma etapa de voo prevista, decolagem em atraso ou, ainda, uma substituição de aeronave. Na base de dados do VRA, estão disponíveis as seguintes informações (ANAC, 2017g):

- **Empresa aérea:** código único que identifica a companhia aérea.
- **Número Voo:** número identificador do voo, dado que está registrado no HOTRAN.
- **Código autorização:** código numérico que permite a identificação do tipo de voo, podendo ser voo regular, voo extra com HOTRAN, voo extra sem HOTRAN, voo de retorno, inclusão de etapa em um voo previsto em HOTRAN, voo cargueiro, voo de serviço, voo de fretamento, voo *charter*, voo de instrução ou voo de experiência.
- **Código tipo de linha:** código que identifica o tipo de linha, podendo ser Cargueiro, Especial, Cargueiro internacional, Sub-Regional, Internacional, Rede Postal, Nacional, Regional.
- **Aeródromo de origem:** código ICAO (valor único por aeródromo) que permite identificar o aeródromo em que ocorreu a decolagem.
- **Aeródromo de destino:** código ICAO (valor único por aeródromo) que permite identificar o aeródromo em que ocorreu o pouso.
- **Partida prevista:** data e hora da partida conforme registrado no HOTRAN.

- **Partida real:** data e hora da partida conforme registro de operação do voo.
- **Chegada prevista:** data e hora da chegada conforme registrado no HOTRAN.
- **Chegada real:** data e hora da chegada conforme registro de operação do voo.
- **Situação voo:** informação que permite classificar se o voo ocorreu ou não.
- **Código justificativa:** código formado por duas letras que permite identificar as justificativas em casos de atrasos e cancelamentos.

Todas as informações de registro das operações aéreas são remetidas pelas companhias aéreas à ANAC, ficando sobre elas a responsabilidade pela autenticidade das informações. Para garantir a veracidade dos dados, a ANAC realiza fiscalização por meio de processos de auditoria. Caso seja encontrado um problema, a ANAC poderá aplicar penalidades administrativas. O procedimento é determinante para a publicação dos dados de VRA, sendo que a ANAC libera a atualização desses dados semanalmente.

#### 2.2.2.5 Tarifas aéreas

Com o estabelecimento da Resolução ANAC n. 437, de 26 de julho de 2017, a Agência passou a disponibilizar dados de tarifas aéreas comercializadas para voos domésticos e internacionais. A disponibilização de tais dados começou em agosto de 2017, porém, são retroativos até o ano de 2002 para voos domésticos e até 2011 para voos internacionais.

No conjunto de informações de tarifas domésticas, são contempladas todas as passagens vendidas ao público adulto em geral, exceto os bilhetes emitidos por meio de programas de milhagens (entre outras condições de gratuidade). Além disso, o número de assentos vendidos possui uma relação que corresponde cerca de 50% do movimento de passageiros pagos transportados.

Para os registros referentes a tarifas de voos internacionais, são aplicadas as mesmas regras de exclusão. Porém, são considerados apenas registros nos quais a origem da viagem é o Brasil e o destino é

no exterior, sendo a viagem integralmente operada pela própria empresa aérea que vendeu as passagens e cujos voos estão autorizados no HOTRAN.

A ANAC, em seu portal, destaca que a base de dados de tarifas foi concebida com propósitos específicos, sendo que a metodologia de apuração dos dados ocasionou diferenças significativas com as demais bases de dados em termos quantitativos. Na base de dados de tarifas, estão disponíveis as seguintes informações (ANAC, 2017e):

- **Ano e mês:** indicam o ano e mês em que as passagens foram vendidas, independentemente da data do voo.
- **Designador OACI da Empresa:** código formado por três letras que indica a companhia aérea que registrou a passagem. Para tarifas aéreas domésticas, o registro é sempre efetuado pela empresa que vendeu a passagem, independentemente da empresa responsável por operar os voos. Para as tarifas aéreas internacionais, somente são registradas as passagens que tenham sido vendidas para voos a serem operados pela própria empresa.
- **Designador OACI de aeroportos (origem e destino):** código formado por quatro letras que indica os aeródromos de origem ou destino do passageiro, independentemente de escalas ou conexão. Nas tarifas aéreas domésticas, passagens de ida e volta e passagens com diversos destinos são fracionadas, sendo apresentadas na forma de passagens de uma origem e um destino. Nas tarifas aéreas internacionais, somente são apontadas passagens as quais o voo se origina no Brasil, podendo ser somente ida ou ida e volta. Desse modo, os dados de tarifas aéreas internacionais contemplam apenas trechos disponíveis no HOTRAN da empresa. Para passagens de ida e volta, o campo "aeródromo de retorno" indica o aeroporto brasileiro do trecho de volta. Para passagens apenas de ida, esse campo é preenchido com o valor 9999.
- **Classes de ida e volta:** este campo é disponibilizado somente para os dados de tarifas aéreas internacionais. É formado por apenas uma letra ou um número, podendo ser:
  - Y: código que indica classe econômica;
  - J: código que indica classe executiva;
  - F: código que indica primeira classe; e
  - 9: código usado para preencher o campo de classe de volta nas passagens apenas de ida.

- **Valor da tarifa:** campo que apresenta o valor do serviço de transporte aéreo constante na passagem. Nessa informação, não estão incluídas as tarifas de embarque dos aeroportos nem os valores cobrados por serviços opcionais. Para as tarifas aéreas domésticas, é apresentado o valor nominal em reais; para as tarifas aéreas internacionais, é apresentado o valor nominal em dólares norte-americanos.
- **Assentos comercializados:** o campo indica o número de assentos consolidados para o ano e mês de venda da passagem, empresa aérea, aeroportos de origem, destino e retorno, classes de cabine de ida e volta e valor de tarifa. Aeroporto de retorno e classes de cabine somente aplicam-se a voos internacionais.

No conjunto de dados relacionados aos voos internacionais, a ANAC cita que inconsistências evidentes foram removidas da base de dados disponibilizada. Em ambos os tipos de voos (domésticos e internacionais), são considerados os códigos OACI dos aeroportos vigentes na data de publicação do dado, pois tais códigos têm alteração ao longo do tempo. Além disso, os dados de tarifas estão em revisão constante, portanto, estão sujeitos a mudança.

### 2.2.3 Empresa Brasileira de Infraestrutura Aeroportuária

A Empresa Brasileira de Infraestrutura Aeroportuária (INFRAERO) é uma empresa pública federal ligada à Secretaria Nacional de Aviação Civil. Iniciou suas atividades de administração, operação e exploração comercial e industrial da infraestrutura aeroportuária brasileira em 1973. Tem como objetivo participar do desenvolvimento do país e do setor aeroportuário, primando pelo bom atendimento aos clientes e pela sustentabilidade do negócio aeroportuário (INFRAERO, 2017b).

Entre as suas responsabilidades, está a gestão de alguns dos aeroportos brasileiros, de unidades de apoio à navegação aérea e de terminais de logísticas de cargas. A principal atribuição da empresa é o desenvolvimento de atividades com o intuito de garantir a prestação dos serviços de segurança, polícia, alfândega e saúde nos aeroportos sob a sua gestão.

A INFRAERO libera, de forma pública em seu portal (<http://www4.infraero.gov.br>), dados resultantes de suas atividades por meio de relatórios. Os dados disponíveis em formato de base de dados são

de estatísticas do setor.

### 2.2.3.1 Estatísticas

A INFRAERO disponibiliza em seu portal estatísticas relacionadas aos aeroportos sob a sua administração. No conjunto de dados disponível, estão os movimentos operacionais de aeronaves, passageiros, carga aérea (somente carregados em porão da aeronave) e correios nos aeroportos. Os dados relacionados aos movimentos militares não estão computados na base de dados estatística.

Os dados de estatísticas disponíveis são datados a partir de janeiro de 2012, em arquivos em formato de planilha eletrônica, consolidados mensalmente. Estão dispostos de forma detalhada por aeroporto e agrupados com o somatório de todos os aeroportos – nesse caso, dados agrupados por toda a administração da INFRAERO. Para cada um dos arquivos mensais, são disponibilizados os dados separados por categoria, sendo elas:

- **Aeronaves:** apresenta dados relacionados a operações de pouso e decolagem de aeronaves. Tais dados são relacionados ao mês de referência do arquivo e consolidados para o ano do arquivo. Para os dados de aeronaves, estão disponíveis as seguintes colunas:
  - Discriminação: coluna que detalha o tipo de voo a que as demais colunas fazem referência, podendo assumir os valores transporte regular (voo doméstico nacional, voo doméstico regional ou voo internacional) ou transporte não regular (voo doméstico ou voo internacional).
  - Pousos no mês: quantidade de operações de pousos realizadas no mês.
  - Pousos no ano: quantidade de operações de pousos realizadas no ano.
  - Decolagens no mês: quantidade de decolagens realizadas no mês.
  - Decolagens no ano: quantidade de decolagens realizadas no ano.
  - Pousos + decolagens no mês: quantidade de pousos somada a operações de decolagens realizadas no mês.
  - Pousos + decolagens no ano: quantidade de pousos somada a operações de decolagens realizadas no ano.

- **Carga aérea:** apresenta dados relacionados a operações de embarque e desembarque de carga aérea. Tais dados são relacionados ao mês de referência do arquivo e consolidados para o ano do arquivo. Para os dados de carga aérea, estão disponíveis as seguintes colunas:
  - Discriminação: coluna que detalha o tipo de voo a que as demais colunas fazem referência, podendo assumir os valores transporte regular (voo doméstico nacional, voo doméstico regional ou voo internacional) ou transporte não regular (voo doméstico ou voo internacional).
  - Carregada no mês: volume de carga aérea em kg carregada nas aeronaves no mês.
  - Carregada no ano: volume de carga aérea em kg carregada nas aeronaves no ano.
  - Descarregada no mês: volume de carga aérea em kg descarregada das aeronaves no mês.
  - Descarregada no ano: volume de carga aérea em kg descarregada das aeronaves no ano.
  - Carga + descarga no mês: volume de carga aérea em kg carregada somada à descarregada no mês.
  - Carga + descarga no ano: volume de carga aérea em kg carregada somada à descarregada no ano.
  
- **Correios:** apresenta dados relacionados a operações de embarque e desembarque de cargas de correios. Tais dados são relacionados ao mês de referência do arquivo e consolidados para o ano do arquivo. Para os dados de correios, estão disponíveis as seguintes colunas:
  - Discriminação: coluna que detalha o tipo de voo a que as demais colunas fazem referência, podendo assumir os valores transporte regular (voo doméstico nacional, voo doméstico regional ou voo internacional) ou transporte não regular (voo doméstico ou voo internacional).
  - Carregada no mês: volume de cargas de correios em kg carregadas nas aeronaves no mês.
  - Carregada no ano: volume de cargas de correios em kg carregadas nas aeronaves no ano.

- Descarregada no mês: volume de cargas de correios em kg descarregadas das aeronaves no mês.
  - Descarregada no ano: volume de cargas de correios em kg descarregadas das aeronaves no ano.
  - Carga + descarga no mês: volume de cargas de correios em kg carregadas somadas às descarregadas no mês.
  - Carga + descarga no ano: volume de cargas de correios em kg carregadas somadas às descarregadas no ano.
- **Passageiros:** apresenta dados relacionados a operações de embarque e desembarque de passageiros. Tais dados são relacionados ao mês de referência do arquivo e consolidados para o ano do arquivo. Para os dados de passageiros, estão disponíveis as seguintes colunas:
    - Discriminação: coluna que detalha o tipo de voo a que as demais colunas fazem referência, podendo assumir os valores transporte regular (voo doméstico nacional, voo doméstico regional ou voo internacional) ou transporte não regular (voo doméstico ou voo internacional).
    - Embarcado no mês: quantidade de passageiros que embarcaram no mês.
    - Embarcado no ano: quantidade de passageiros que embarcaram no ano.
    - Desembarcado no mês: quantidade de passageiros que desembarcaram no mês.
    - Desembarcado no ano: quantidade de passageiros que desembarcaram no ano.
    - Embarcado + desembarcado no mês: quantidade de passageiros que embarcaram somada aos que desembarcaram no mês.
    - Embarcado + desembarcado no ano: quantidade de passageiros que embarcaram somada aos que desembarcaram no ano.

Para o entendimento dos dados, segundo a INFRAERO (2017a), é necessário atentar para os conceitos relacionados ao tipo de transporte e às classificações de passageiros, a saber:

- **Transporte regular:** todos os voos efetuados com registro de HOTRAN.

- **Transporte não regular:** todos os voos comerciais efetuados por qualquer aeronave e que possuem registro de HOTRAN.
- **Passageiros embarcados:** quantidade de passageiros de origem (que iniciam a viagem no aeroporto) somada à quantidade de passageiros em conexão.
- **Passageiros desembarcados:** quantidade passageiros de destino (que terminam a viagem no aeroporto) somada à quantidade de passageiros em conexão.

A INFRAERO (2017a) informa que o grau de confiabilidade dos dados pode estar prejudicado pela publicação da Resolução n. 8/A-NAC/2007. No conjunto de ações de tal Resolução, está a revogação da obrigatoriedade do repasse das informações de passageiros embarcados, sendo que essa demanda havia sido estabelecida pela Portaria n. 602/GC5/2000 do Comando da Aeronáutica. Ou seja, as informações disponíveis podem divergir da realidade.

## 2.2.4 Departamento de Controle do Espaço Aéreo

O Departamento de Controle do Espaço Aérea (DECEA) é uma organização com a atribuição de controlar o espaço aéreo brasileiro. O DECEA provê os serviços de navegação aérea com a finalidade de viabilizar os voos e ordenar os fluxos de tráfego aéreo no Brasil (DECEA, 2017).

Outra atribuição do DECEA é a gestão do Sistema de Controle do Espaço Aéreo Brasileiro (SISCEAB). O órgão estabelece metas e atribuições a serem cumpridas pelas organizações, subordinadas, que são responsáveis pela execução operacional do SISCEAB.

Para que o Departamento possa executar e gerir suas atividades, incorpora recursos humanos especializados e detém *expertise* e tecnologias indispensáveis para a execução dos complexos procedimentos referentes às estratégias do SISCEAB. Sua estrutura física é robusta e está presente em uma centena de municípios de todas as 27 unidades federativas brasileiras.

O DECEA libera, de forma pública, dados resultantes das suas atividades de controle do espaço aéreo, os quais estão disponíveis no portal do Departamento (<https://www.decea.gov.br>).

### 2.2.4.1 Serviço de Informação Aeronáutica

O Serviço de Informação Aeronáutica (AISWEB) é a fonte oficial de informações aeronáuticas produzidas pelo DECEA em meio digital. Estão disponíveis avisos aos aeronavegantes, cartas aeronáuticas, informações meteorológicas e publicações acerca dos aeródromos.

As informações estão disponíveis de duas formas: diretamente pelo portal do Departamento ou por conjuntos de serviços *web* denominados API (*Application Programming Interface*). No contexto deste trabalho, serão exploradas as opções disponíveis no API, pois são passíveis de consultas diretamente computador a computador.

Para consumir informações da API do AISWEB, inicialmente deve ser realizado um cadastro no Portal do DECEA ([www.aisweb.aer.mil.br/\\_inc/iframe.cfm?pagina=api](http://www.aisweb.aer.mil.br/_inc/iframe.cfm?pagina=api)), sendo os dados de acesso liberados por *e-mail* (KEY e PASS). Após o recebimento dos dados de acesso, deve-se se acessar a API por meio de uma URL com a seguinte estrutura básica: [http://www.aisweb.aer.mil.br/api/?apiKey=\[APIKEY\]&apiPass=\[APIPass\]&area=\[API\]](http://www.aisweb.aer.mil.br/api/?apiKey=[APIKEY]&apiPass=[APIPass]&area=[API]). O retorno da API do DECEA é um conjunto de dados em formato XML, com um *layout* específico para cada conjunto de informações. Os parâmetros que configuram quais informações serão retornadas pela API são os seguintes:

- **API KEY:** neste campo, deve ser passada a informação de chave da API que é enviada por *e-mail*.
- **API Pass:** neste campo, deve ser passado o código único da API enviada por *e-mail*.
- **API:** neste campo, deve ser passado o tipo de informação que se deseja consumir, podendo ser:
  - **Aeródromo:** disponibiliza informações sobre o aeródromo.
  - **Cartas:** disponibiliza informações de cartas aeronáuticas.
  - **NOTAM** (*Notice to Airmen*): disponibiliza avisos aos aeronavegantes.
  - **Suplementos de AIP (Publicação de Informação Aeronáutica):** disponibiliza informações de natureza explicativa, de assessoramento e até mesmo administrativa ou técnica sobre o aeródromo.
  - **Nascer e pôr do sol:** disponibiliza informações de horário do nascer e do pôr do sol no aeródromo.

- **METAR** (Informe meteorológico regular de aeródromo): disponibiliza informações de meteorologia na região do aeródromo de forma codificada.

O serviço referente aos dados de aeródromos retorna informações de nome do aeródromo, cidade, UF, latitude, longitude, condições de voos e características meteorológicas da região em que o aeródromo se encontra. Tais informações estão disponíveis por meio da inclusão dos seguintes parâmetros na URL (&area=[CODIGO]&IcaoCode=[ICAO]):

- **area**: código da API que caracteriza o tipo de informação resultante; nesse caso, deve ser passado o código "aero".
- **IcaoCode**: código ICAO que identifica o aeródromo sobre o qual se deseja obter informações.

O serviço referente às cartas traz informações de cartas aeronáuticas da região de determinado aeródromo, rota e situações visuais de voo em vigor. Tais informações estão disponíveis por meio da inclusão dos seguintes parâmetros na URL (&area=[CODIGO]&IcaoCode=[ICAO]&especie=[CODIGO]&tipo=[CODIGO]):

- **area**: código da API que caracteriza o tipo de informação resultante; nesse caso, deve ser passado o código "cartas".
- **IcaoCode**: código ICAO que identifica o aeródromo sobre o qual se deseja obter informações.
- **especie**: código que permite definir o tipo de operação sobre o qual se deseja as informações, podendo ser IFR (aeródromos e rota) ou VFR (visuais).
- **tipo**: código que permite especificar o tipo de carta desejada, podendo assumir:
  - ADC: código para visualização de cartas de aeródromo.
  - AOC: código para visualização de cartas de obstáculo.
  - ARC: código para visualização de carta de área.
  - ATCSMAC: código para visualização de cartas de altitude mínima de vigilância.
  - GMC: código para visualização de cartas de movimento no solo.

- IAC: código para visualização de cartas de aproximação por instrumentos.
- LC: código para visualização de cartas de pouso.
- PATC: código para visualização de cartas topográficas de aproximação de precisão.
- PDC: código para visualização de cartas de estacionamento de aeródromos.
- SID: código para visualização de cartas de saída normalizada – voo por instrumentos.
- STAR: código para visualização de cartas de chegada normalizada – voo por instrumentos.
- VAC: código para visualização de cartas de aproximação visual.
- OTR: código para visualização de outras cartas.

O serviço referente ao NOTAM traz informações de vários aeródromos ou de um FIR (Região de Informação de Voo). Existem informações relativas a estabelecimento, condição ou modificação de qualquer instalação aeronáutica, serviço, procedimento ou perigo. Tais conhecimentos oportunos são indispensáveis para o pessoal encarregado das operações de voo. As informações estão disponíveis por meio da inclusão dos seguintes parâmetros na URL (&area=[CODIGO]&IcaoCode=[LISTA\_ICAO]&dist=[CODIGO]):

- **area:** código da API que caracteriza o tipo de informação resultante; nesse caso, deve ser passado o código "notam".
- **IcaoCode:** listagem de códigos identificadores do aeródromo ou da região desejada, separados por vírgula.
- **Dist:** código identificador do tipo de NOTAM; nesse caso, deve ser passado o parâmetro "N" para Nacional ou "I" para Internacional.

O serviço referente a suplementos de publicação de informação aeronáutica traz informações de determinado aeródromo, como título da informação, texto (descritivo), duração e anexo (podendo ser um documento com detalhes da informação). Tais informações estão disponíveis por meio da inclusão dos seguintes parâmetros na URL (&area=[CODIGO]&IcaoCode=[ICAO]):

- **area:** código da API que caracteriza o tipo de informação resultante; nesse caso, deve ser passado o código "suplementos".
- **IcaoCode:** código identificador do aeródromo ou da região desejada.

O serviço referente aos dados de nascer e pôr do sol traz informações de horários do nascer e do pôr do sol para determinado aeródromo para as datas utilizadas como parâmetro. Tais informações estão disponíveis por meio da inclusão dos seguintes parâmetros na URL (&area=[CODIGO]&IcaoCode=[ICAO]&dt\_i=[AAAMMDD]&dt\_f=[AAAMMDD]):

- **area:** código da API que caracteriza o tipo de informação resultante; nesse caso, deve ser passado o código "sol".
- **IcaoCode:** código identificador do aeródromo.
- **dt\_i:** data de início da consulta formada por ano, mês e dia.
- **dt\_f:** data de fim da consulta formada por ano, mês e dia.

O serviço referente aos Dados de informe meteorológico regular de aeródromo traz informações de meteorologia na região de determinado aeródromo e do terminal, bem como localidade. Tais informações estão disponíveis por meio da inclusão dos seguintes parâmetros na URL (&area=[CODIGO]&IcaoCode=[ICAO]):

- **area:** código da API que caracteriza o tipo de informação resultante; nesse caso, deve ser passado o código "metar".
- **IcaoCode:** código identificador do aeródromo.

O DECEA informa que a API do AISWEB está liberada e é de uso irrestrito, incluindo fins comerciais, porém, como mencionado, mediante registro de instituição, nome do requerente, *e-mail* e aplicação que irá utilizar. Tais dados, por estarem disponíveis em formato XML, são passíveis de manipulação por computadores e viabilizam a sua integração com outros sistemas.

### 2.2.5 Secretaria de Aviação Civil do Ministério dos Transportes, Portos e Aviação Civil

A Secretaria Aviação Civil do Ministério dos Transportes, Portos e Aviação Civil (SAC/MTPA) foi criada em 2011 pela Lei n. 12462/2011

com a finalidade de coordenar e supervisionar ações voltadas ao desenvolvimento estratégico do setor da aviação civil e da infraestrutura aeroportuária e aeronáutica no Brasil (SAC, 2017b).

A SAC libera, de forma pública, dados resultantes de tais atividades, os quais estão disponíveis no portal da Secretaria (<http://www.aviacao.gov.br>) para o acesso de qualquer cidadão. Entre eles, podem ser citados dados de cadastro de aeródromos, dados estatísticos, dados de horário de transporte e pesquisa de satisfação de passageiros.

### 2.2.5.1 Sistema Hórus

Para a disponibilização de informações do setor, a SAC/MTPA apresentou, em março de 2017, o sistema Hórus, uma base de dados com informações de 303 aeródromos públicos e privados. Nele, estão disponíveis informações de Movimentação (dados oriundos da ANAC), dados de cadastro de aeródromos (dados oriundos da ANAC), HOTRAN (dados oriundos da ANAC) e Desempenho Operacional (dados da própria SAC) (SAC, 2017c).

Pelo fato de os dados de Movimentação, Cadastro de Aeródromos e HOTRAN já terem sido contemplados neste trabalho por meio da exploração das informações da ANAC, na identificação das informações disponibilizadas pela SAC/MTPA, será tratada apenas a parte de Desempenho Operacional.

Os dados de Desempenho Operacional são resultados de uma pesquisa compilada trimestralmente acerca da satisfação dos usuários do transporte aéreo no que diz respeito a 38 itens de infraestrutura, atendimento, serviços e itens de gestão. A pesquisa é realizada de forma permanente em 15 aeroportos do Brasil, e seu objetivo é mapear e monitorar o desempenho operacional dos terminais e a percepção do usuário sobre o serviço, além de oferecer bases para o planejamento de políticas públicas para o setor de aviação civil (SAC, 2017a).

No sistema Hórus, estão disponíveis dados de Desempenho Operacional datados a partir de janeiro de 2013. No contexto desse sistema, todas as informações são disponibilizadas por meio de gráficos, infográficos e grades de dados, além de existir a opção de exportar tais informações em formato CSV (por gráfico) ou por meio das interfaces de dados completos a base de dados completa. Para os dados de Desempenho Operacional, estão disponíveis as seguintes informações (SAC, 2017c):

- **Aeródromo:** nome do aeródromo da pesquisa.

- **ICAO:** código identificador do aeródromo.
- **Processo:** processo em que o passageiro se encontra no momento da entrevista, podendo ser embarque ou desembarque.
- **Tipo de voo:** tipo de voo em que o passageiro estava viajando, podendo ser doméstico ou internacional.
- **Terminal:** número do terminal em que a entrevista foi realizada.
- **Data da Coleta:** data em que a entrevista foi realizada.
- **Mês da coleta:** mês em que a entrevista foi realizada.
- **Horário do início da coleta:** horário de início da entrevista.
- **Horário do fim da coleta:** horário de término da entrevista.
- **Forma de *check-in*:** modo de *check-in* realizado pelo passageiro entrevistado.
- **Tipo de transporte para o aeroporto:** tipo de transporte usado pelo passageiro para deslocar-se até o aeroporto, podendo ser privado, público ou em conexão.
- **Meio de transporte público utilizado:** descreve o tipo de transporte público utilizado até o aeroporto.
- **Disponibilidade de carrinhos de bagagem:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Facilidade de embarque/desembarque no meio-fio:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo na fila da inspeção de segurança (Raio-X):** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Rigor da inspeção de segurança:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Cordialidade e atendimento dos funcionários da inspeção de segurança (Raio-x):** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.

- **Qualidade da sinalização do aeroporto:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Disponibilidade de painéis de informações de voos:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Distância caminhada no terminal de passageiros:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Cordialidade dos funcionários do aeroporto:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Disponibilidade de tomadas:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Qualidade da internet disponibilizada pelo aeroporto:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Disponibilidade de sanitários:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Limpeza dos sanitários:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Disponibilidade de assentos da sala de embarque:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Sensação de segurança nas áreas públicas do aeroporto:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Limpeza geral do aeroporto:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.

- **Conforto na sala de embarque:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Conforto térmico do aeroporto:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Conforto acústico do aeroporto:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Identificação do número do voo na esteira de restituição de bagagem:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Qualidade das instalações de estacionamento de veículos:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Custo-benefício do estacionamento:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Quantidade e qualidade de lanchonetes e restaurantes:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Custo-benefício dos produtos de lanchonetes/restaurantes:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo de fila em lanchonetes/restaurantes:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Cordialidade dos funcionários de lanchonetes/restaurantes:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Disponibilidade e localização de bancos, caixas eletrônicos e casas de câmbio:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.

- **Quantidade e qualidade dos estabelecimentos comerciais:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Custo-benefício dos produtos comerciais:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo de fila nos estabelecimentos comerciais:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Cordialidade dos funcionários do comércio:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Qualidade sala VIP:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo na fila do *check-in* eletrônico:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo na fila do *check-in* balcão:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Eficiência dos funcionários do *check-in*:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Cordialidade e atendimento dos funcionários do *check-in* balcão:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Facilidade para realizar conexões:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo de espera na restituição de bagagem:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.

- **Integridade da bagagem após a restituição:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo na fila do controle de emigração:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Cordialidade e atendimento dos funcionários do controle de emigração:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo de fila na imigração:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Cordialidade dos funcionários da imigração:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Tempo na fila do controle aduaneiro:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Cordialidade dos funcionários do controle aduaneiro:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Disponibilidade de transporte público para o aeroporto:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Disponibilidade de táxi:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Qualidade da informação prestada pela companhia aérea:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Disponibilidade de vagas no estacionamento de veículo:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.

- **Confiabilidade da inspeção de segurança (Raio-X):** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Satisfação geral:** indicador quantitativo relacionado à satisfação do passageiro entrevistado quanto ao item, podendo assumir valores de 0 a 5.
- **Nacionalidade do entrevistado:** indica a nacionalidade do passageiro entrevistado.
- **Gênero do entrevistado:** indica o gênero do passageiro entrevistado.
- **Idade do entrevistado:** indica a idade do passageiro entrevistado.
- **Escolaridade do entrevistado:** indica a escolaridade do passageiro entrevistado.
- **Profissão do entrevistado:** indica a profissão do passageiro entrevistado.
- **Renda do entrevistado:** indica a renda do passageiro entrevistado em faixas de valores.
- **Viajando sozinho:** indica se o passageiro está viajando sozinho ou não, podendo assumir os valores sim ou não.
- **Número de acompanhantes:** indica o número de acompanhantes do passageiro, podendo assumir 1, 2, 3, 4 ou mais.
- **Motivo de viagem:** indica o motivo da viagem do passageiro, podendo assumir lazer, negócios ou outros.
- **Companhia aérea:** indica o nome da companhia aérea em que está sendo realizada a viagem.
- **Número do voo:** indica o número do voo do passageiro.
- **Quantidade de viagens partindo deste aeroporto:** indica a quantidade de viagens do passageiro partindo deste aeroporto (da entrevista).
- **Frequência de viagem:** indica a frequência de viagem do passageiro, podendo assumir 0 a 2, 3 a 4 ou mais.

- **Tempo de antecedência de chegada para o voo:** indica o tempo de antecedência de chegada do passageiro antes do voo.

Para o conjunto de dados, a SAC/MTPA classifica os indicadores em dois grupos quantitativos e qualitativos. Os indicadores quantitativos são os dados que o passageiro informa por meio de notas de 0 a 5; com elas, são geradas médias aritméticas simples. Já os indicadores qualitativos são os dados que o passageiro repassa de forma "aberta", sem uma nota diretamente atribuída ao indicador; nesse caso, os dados são disponibilizados em formato de proporção.

Os dados de Desempenho Operacional são divulgados pela SAC/MTPA de forma compilada para cada trimestre/ano (média aritmética de cada indicador), isso por meio de relatórios e sistema Hórus. Os dados disponibilizados de forma completa estão detalhados por entrevista.

### **2.2.6 Ministério da Indústria, Comércio Exterior e Serviços**

O Ministério da Indústria, Comércio Exterior e Serviços (MDIC) é um órgão integrante da estrutura da administração pública federal direta que tem como objetivos estabelecer, cumprir e avaliar políticas públicas para a promoção da competitividade, do comércio exterior, do investimento e da inovação nas empresas e do bem-estar do consumidor (MDIC, 2017a).

O MDIC é referência na gestão de políticas públicas para fortalecer a competitividade das empresas brasileiras. Para tanto, é vinculado às seguintes entidades: Superintendência da Zona Franca de Manaus (SUFRAMA), Instituto Nacional da Propriedade Industrial (INPI) e Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO), os quais apoiam ativamente as atividades do MDIC.

Além das entidades públicas citadas, junto ao MDIC está a Agência Brasileira de Desenvolvimento Industrial (ABDI), que é uma entidade privada e sem fins lucrativos responsável, na celebração do Contrato de Gestão do MDIC, pelos recursos para a realização de ações de interesse público.

O MDIC libera dados resultantes de suas atividades relacionadas ao estabelecimento de políticas públicas do comércio exterior, os quais estão disponíveis no portal do Ministério (<https://www.mdic.gov.br>) para o acesso de qualquer cidadão. Considerando o contexto deste trabalho, o MDIC disponibiliza dados de produtos exportados e importados que são transportados por via aérea.

### 2.2.6.1 Base de dados do comércio exterior brasileiro

A base de dados do comércio exterior brasileiro é a referência do MDIC para a construção da balança comercial brasileira. Além disso, tal base é a fonte de dados para o Sistema de Análise das Informações de Comércio Exterior (Alice Web), responsável por divulgar as estatísticas brasileiras de exportações e importações (MDIC, 2017b, 2017a).

O conjunto de dados disponível na base de dados de comércio exterior permite caracterizar e avaliar o movimento comercial do Brasil com os demais países, apresentando informações de exportações e importações efetuadas pelo país. Por apresentar dados históricos acerca das mercadorias comercializadas, permite identificar o desempenho da economia brasileira, constituindo-se como um instrumento básico para a tomada de decisão econômica por parte do governo brasileiro (MDIC, 2017b).

A base de comércio exterior é disponibilizada em formato CSV, de forma detalhada por NCM (Nomenclatura Comum do Mercosul) e por municípios. Além disso, para cada tipo de disponibilização (NCM ou municípios), constam informações de exportação e importação de forma separada, isso a partir de janeiro 1997. No contexto do MDIC, exportação corresponde a mercadorias embarcadas para o exterior, sem retorno previsto; já importação corresponde à entrada de mercadorias originárias do exterior, sem retorno previsto.

De forma detalhada por NCM, as bases de dados de exportação e importação disponibilizam as seguintes informações (MDIC, 2017b, 2017a):

- **CO\_ANO:** corresponde ao ano de referência do dado.
- **CO\_MES:** corresponde ao mês de referência do dado.
- **CO\_NCM:** corresponde ao código identificador do NCM da mercadoria exportada ou importada. Nessa base de dados, o código NCM é composto de oito dígitos, sendo os seis primeiros formados pelo Sistema Harmonizado (Capítulo, Posição e Subposição) e os dois últimos (item e subitem) criados de acordo com a definição estabelecida entre os países do Mercosul.
- **CO\_UNID:** corresponde ao código identificador da unidade estatística.
- **CO\_PAIS:** corresponde ao código identificador do país de origem (importação) ou destino (exportação).

- **CO\_UF:** corresponde ao código identificador do estado produtor (exportação) ou importador (importação).
- **CO\_PORTO:** corresponde ao código identificador da unidade onde ocorreu o embarque da mercadoria (exportação) ou o desembarque da mercadoria (importação).
- **CO\_VIA:** corresponde ao código identificador da via de transporte utilizada para o transporte da mercadoria. Na exportação, configura-se o último meio de transporte utilizado para movimentar a mercadoria até seu destino; já na importação, configura-se o primeiro meio de transporte da mercadoria no Brasil.
- **QT\_ESTAT:** corresponde à quantidade estatística da movimentação da mercadoria.
- **KG\_LIQUIDO:** corresponde à quantidade em quilogramas líquidos da mercadoria movimentada.
- **VL\_FOB:** corresponde ao valor FOB (*Free On Board*), expresso em dólares americanos, referente à movimentação da mercadoria.

De forma detalhada por municípios, as bases de dados de exportação e importação disponibilizam as seguintes informações (MDIC, 2017b, 2017a):

- **CO\_ANO:** corresponde ao ano de referência do dado.
- **CO\_MES:** corresponde ao mês de referência do dado.
- **SH4:** corresponde ao código identificador da mercadoria exportada ou importada. Nessa base de dados, o código NCM é composto de quatro dígitos, formados pelo Sistema Harmonizado (Capítulo, Posição e Subposição), estando de acordo com a definição estabelecida entre os países do Mercosul.
- **CO\_PAIS:** corresponde ao código identificador do país de origem (importação) ou destino (exportação).
- **CO\_UF:** corresponde ao código identificador do estado produtor (exportação) ou importador (importação).
- **CO\_PORTO:** corresponde ao código identificador da unidade onde ocorreu o embarque da mercadoria (exportação) ou o desembarque da mercadoria (importação).

- **CO\_MUN\_GEO:** corresponde ao código identificador dos municípios-sede (domicílio fiscal) das empresas que realizaram as operações de exportação e importação.
- **KG\_LIQUIDO:** corresponde à quantidade em quilogramas líquidos da mercadoria movimentada.
- **VL\_FOB:** corresponde ao valor FOB (*Free On Board*), expresso em dólares americanos, referente à movimentação da mercadoria.

Ambos os formatos de bases de dados disponibilizam o detalhamento de ano, mês, país, estado, porto, quantidade em quilogramas líquido e valor FOB. Entretanto, na base de dados por NCM, as mercadorias contam com detalhamento por via de transporte. Já na base de dados de municípios, as mercadorias não apresentam via de transporte.

Na base por NCM, as vias de transportes disponíveis são marítima, fluvial, lacustre, aérea, postal, ferroviária, rodoviária, tubo-conduto, linha de transmissão e meios próprios. No contexto deste trabalho, é importante considerar os dados referentes à via aérea. Para a identificação dos aeroportos, deve ser considerado o campo porto, porém, este não é especificado por código ICAO, como ocorre nas demais bases de dados analisadas. Assim, devem ser considerados os portos cujo sufixo é "Aeroporto". Isso posto, constam informações de 34 aeroportos para ambos os formatos de bases disponibilizadas.

As bases de dados de comércio exterior são disponibilizadas mensalmente, sendo liberadas de modo público e atualizadas com os dados do mês mais recente entre três e 10 dias após o final do referido mês, e têm como fonte as informações o Sistema Integrado de Comércio Exterior (SISCOMEX).

### 2.2.7 Resumo

Na tabela 1 é apresentado um resumo dos dados identificados na seção 2.2.

## 2.3 INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS

Conforme apresentado em MPOG e CONCAR (2010), o aprimoramento das áreas de ciência da computação e tecnologia da informação quanto ao desenvolvimento de dados espaciais tem viabilizado a popularização e o uso de geotecnologias, tais como Sistemas de Informações

Geográficas, Sensoriamento Remoto, Sistemas de Posicionamento Global e Serviços Baseados em Localização.

A integração entre dados geoespaciais de forma consistente requer conhecimento acerca de conceitos, normas e especificações inerentes aos dados e aos usos aos quais eles serão destinados. MPOG e CONCAR (2010) definem bases geográficas ou geoespaciais como conjuntos de dados que possuem duas dimensões: uma quanto à identificação e à localização na superfície da Terra (relação a um sistema geodésico de referência) e outra quanto ao seu conjunto de informações descritivas.

Segundo MPOG e CONCAR (2011), existe um grande número de instituições que estão envolvidas na produção e na distribuição de dados espaciais. É importante que as instituições produtoras de tais dados mantenham aderência a um conjunto de padrões e normas. Essa ação visa garantir a interoperabilidade entre diversos sistemas e o compartilhamento das informações entre as organizações envolvidas. Portanto, faz-se necessária a existência de padrões de metadados consolidados e estruturados, garantindo os seguintes aspectos:

- identificação do produtor e da responsabilidade técnica de produção;
- especificação das terminologias utilizadas;
- garantia da transferência de dados;
- viabilização da integração de informações;
- identificação da qualidade das informações geográficas, subsidiando a análise do usuário quanto à adequação a suas aplicações;
- garantia dos requisitos mínimos de divulgação e uso dos dados geoespaciais.

De acordo com Paixão, Nichols e Coleman (2008), a expressão "Infraestrutura de Dados Espaciais" (IDE) surgiu com a finalidade de relacionar fontes de dados, sistemas, redes, padrões e questões governamentais a respeito de informações e dados espaciais que são entregues aos potenciais usuários por diferentes meios. Para os autores, uma IDE deve ainda facilitar a busca do usuário por dados espaciais.

Conforme descrito em MPOG e CONCAR (2010), o fator que acelerou o desenvolvimento das IDEs no mundo foi a conscientização da sociedade sobre a importância acerca do compartilhamento de bases de dados geoespaciais com a finalidade de integração, compatibilização e disponibilização das informações consideradas de uso comum. No Brasil, por meio do Decreto n. 6.666, de 27 de novembro de 2008,

foi estabelecida a Infraestrutura Nacional de Dados Espaciais (INDE), caracterizada como:

”conjunto integrado de tecnologias; políticas; mecanismos e procedimentos de coordenação e monitoramento; padrões e acordos, necessário para facilitar e ordenar a geração, o armazenamento, o acesso, o compartilhamento, a disseminação e o uso dos dados geoespaciais de origem federal, estadual, distrital e municipal” (BRASIL, 2008, p. 1).

Conforme apresentado em BRASIL (2008), a INDE tem como objetivo promover o adequado ordenamento na geração, no armazenamento, no acesso, no compartilhamento, na disseminação e no uso dos dados geoespaciais; promover a utilização dos dados geoespaciais; e evitar a duplicidade de ações e o desperdício de recursos na obtenção de dados geoespaciais pelos órgãos da administração pública por meio da divulgação dos metadados relativos a esses dados disponíveis nas entidades. Nesse sentido, a INDE tem a finalidade de prover subsídios para um melhor proveito das informações geoespaciais disponíveis para a sociedade.

Para garantir a atendimento aos objetivos da INDE, foi estabelecido um conjunto de atores que determinam os propósitos em uma IDE. Segundo CONCAR (2009), os atores propostos para a INDE são:

- **Organizacionais e administrativos:** são responsáveis pelo planejamento do orçamento, pela formulação de planos e prioridades, pelo estabelecimento de regras e responsabilidades, fomento à participação e apoio à elaboração de acordos de cooperação e de compartilhamento de dados.
- **Produtores de dados e informações geoespaciais de referência e temáticas:** são responsáveis pela definição e identificação das informações geoespaciais que devem ser consideradas de referência e temáticas na INDE. Os atores produtores das informações são de esfera federal, sendo cada órgão responsável por uma ou mais temáticas.
- **Usuários:** são responsáveis pela produção das informações; tratam-se de entes governamentais que utilizarão a informação para a formulação de políticas públicas baseadas na dimensão geoespacial. Ainda, tais entes também poderão beneficiar-se, como usuários,

de todo o volume de informações geográficas a ser disponibilizado na INDE, retornando valor para a sociedade por meio de melhores produtos e serviços.

- **Produtores de dados e informações de valor agregado:** são responsáveis pela disponibilização de dados de referência e temáticos por determinado interesse e utilização específica; que podem pertencer aos âmbitos setoriais, regionais, estaduais, municipais, urbanos e outros.
- **Provedores de produtos e serviços nas áreas de geoprocessamento e correlatas:** atores que compõem os setores de serviços de geoprocessamento, geomática, geotecnologias e tecnologias da geoinformação em expansão no Brasil. Empresas do setor privado fornecedoras de produtos de *hardware* e *software*, bem como de serviços de projeto e desenvolvimento de sistemas/aplicações, projeto e construção de bases de dados geoespaciais, suporte operacional em sistemas, treinamento e consultoria na área.

A INDE é composta por vários atores e, para que seus objetivos sejam alcançados, foi disponibilizado pelo Ministério do Planejamento, Orçamento e Gestão e pela Comissão Nacional de Cartografia um documento denominado "Plano de Ação para Implantação da Infraestrutura Nacional de Dados Espaciais". O referido documento aborda questões conceituais da INDE, diretrizes para sua implantação, questões organizacionais, técnicas e humanas. Além do Plano de Ação, no contexto da INDE, existem outros documentos importantes, os quais apresentam especificações técnicas acerca dos dados espaciais, tais como:

- **ET-EDGV – Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais:** o documento apresenta definições para um modelo conceitual acerca dos dados geoespaciais vetoriais.
- **ET-EDGV-Defesa-F-Ter – Especificação Técnica para Estruturação de Dados Geoespaciais Vetoriais de Defesa da Força Terrestre:** o documento apresenta definições para um modelo conceitual acerca dos dados geoespaciais vetoriais que estão relacionados às atividades de planejamento e execução de ações de defesa e segurança no Espaço Geográfico Brasileiro (EGB), a serem desenvolvidas pela Força Terrestre (Exército Brasileiro).

- **ET-ADGV – Especificação Técnica para a Aquisição de Dados Geoespaciais Vetoriais:** o documento apresenta definições de regras para aquisição da geometria dos dados.
- **ET-PCDG – Especificação Técnica de Produtos de Conjuntos de Dados Geoespaciais:** o documento apresenta definições dos padrões dos produtos vetoriais e matriciais.
- **ET-RDG – Especificação para a Representação de Dados Geoespaciais:** o documento apresenta definições para garantir a consistência na representação das classes de objetos.
- **ET-CQDG – Especificação Técnica para o Controle de Qualidade dos Produtos de Conjuntos de Dados Geoespaciais:** o documento apresenta definições de procedimentos para o controle de qualidade dos produtos.
- **PerfilMGB – Perfil de Metadados Geoespaciais do Brasil:** o documento apresenta definições do conjunto estruturado de elementos básicos e necessários que retratam as características dos produtos geoespaciais brasileiros, garantindo sua identificação, sua avaliação e sua utilização consistentes.

O ET-EDGV-Defesa-F-Ter apresenta uma proposição de estrutura de dados para cada um dos sistemas de transportes. Considerando o contexto deste trabalho, o documento traz uma proposição para o sistema de transportes aeroportuário. Na Figura 8, é possível visualizar um modelo conceitual para o sistema de transporte aeroportuário, no qual um complexo aeroportuário pode conter uma ou mais pistas. No ET-EDGV-Defesa-F-Ter, um complexo aeroportuário é definido como:

”[...] um conjunto de elementos físicos cuja finalidade é apoiar as atividades relacionadas ao sistema portuário, devendo o mesmo ser homologado pela ANAC” (MD, 2016, p. 114).

A estrutura de dados para um complexo aeroportuário é formada por sete atributos, sendo eles:

- **Sigla ICAO:** informação alfanumérica de quatro caracteres que indica o código do complexo aeroportuário estabelecido pela ICAO e constante no ROTAER.
- **Sigla IATA:** informação alfanumérica de três caracteres que indica a sigla oficial do complexo aeroportuário estabelecido pela IATA.

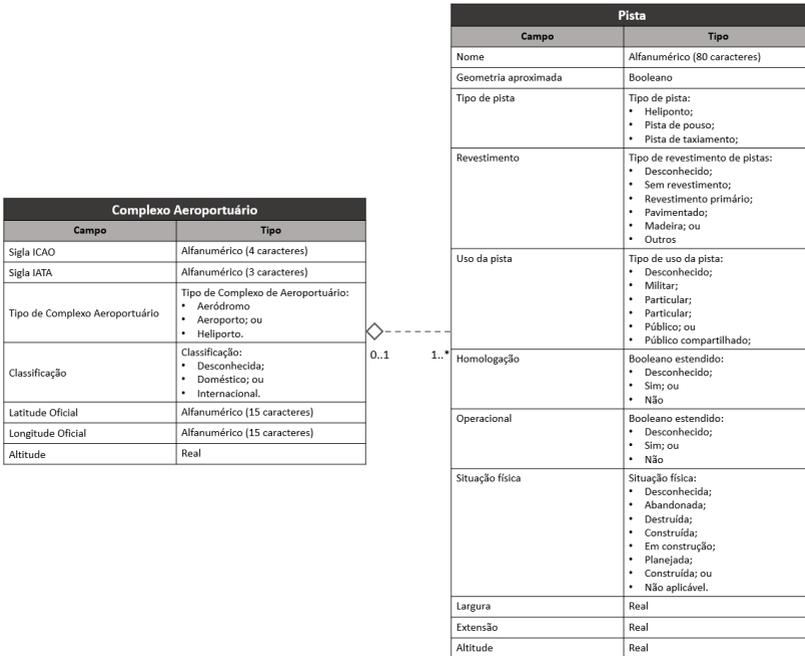


Figura 8 – Estrutura de dados proposta pela INDE para complexo aeroportuário.

Fonte: Adaptada de MD (2016).

- Tipo de Complexo Aeroportuário:** informação alfanumérica oriunda da classe Tipo de Complexo de Aeroportuário e que pode assumir os seguintes valores: aeroporto (todo aeródromo público dotado de instalações e facilidades para apoio a operações de aeronaves, embarque e desembarque de pessoas e cargas); aeródromo (toda área destinada a pouso, decolagem e movimentação de aeronaves); ou heliponto (heliponto público dotado de instalações e facilidades para o apoio a operações de helicópteros e embarque e desembarque de passageiros e cargas). Esse campo identifica o tipo de complexo aeroportuário.
- Classificação:** informação alfanumérica oriunda da classe Classificação e que pode assumir os seguintes valores: desconhecida (valor desconhecido); doméstico (destina-se somente a serviços domésticos de aeronaves); ou internacional (destina-se a serviços

internacionais e domésticos, se for o caso, de aeronaves). Esse campo indica a classificação do complexo aeroportuário quanto à operação.

- **Latitude oficial:** informação alfanumérica de 15 caracteres que indica a latitude oficial (ROTAER) do ponto de referência do complexo.
- **Longitude oficial:** informação alfanumérica de 15 caracteres que indica a longitude oficial (ROTAER) do ponto de referência do complexo.
- **Altitude:** informação do tipo numérica real que indica a altitude oficial (ROTAER) do ponto de referência do complexo (em metros).

A composição de metadados para pistas é formada por sete atributos, sendo eles:

- **Nome:** informação alfanumérica de 80 caracteres que indica o nome completo da instância.
- **Geometria Aproximada:** informação do tipo booleano que indica que a geometria adquirida é aproximada em relação à escala prevista para o produto cartográfico.
- **Tipo de Pista:** informação alfanumérica que identifica o tipo de pista, oriunda da classe Tipo de Pista, podendo assumir os seguintes valores: heliponto (ponto ou área situada ao nível do solo ou acima dele, destinada a pousos e decolagens de helicópteros); pista de pouso (área plana designada a pousos e decolagens de aeronaves); ou pista de taxiamento (pistas de acesso interno de aeroportos, destinadas a aeronaves).
- **Revestimento:** informação alfanumérica que identifica a natureza do revestimento da pista, oriunda da classe Revestimento, podendo assumir os seguintes valores: desconhecido (valor desconhecido); sem revestimento (próprio terreno natural); revestimento primário (saibro e/ou cascalho); pavimentado (revestimento rígido ou flexível, como asfalto ou concreto, asfáltico, concreto, cimento, paralelepípedos, pedras irregulares); madeira ou outros (outros tipos de revestimento).

- **Uso da pista:** informação alfanumérica que indica o responsável pela administração da pista ou do ponto de pouso. Os possíveis valores são oriundos da classe Uso Principal, podendo ser: desconhecido; militar; particular; público; ou público compartilhado.
- **Homologação:** informação do tipo booleano (estendido) que indica se a pista ou o ponto de pouso é homologado ou não. Os possíveis valores são oriundos da classe Booleano Estendido, podendo ser: desconhecido (valor desconhecido); sim (valor booleano verdadeiro); ou não (valor booleano falso).
- **Operacional:** informação do tipo booleano (estendido) que indica se a pista ou o ponto de pouso está operacional. Os possíveis valores são oriundos da classe Booleano Estendido, podendo ser: desconhecido (valor desconhecido); sim (valor booleano verdadeiro); ou não (valor booleano falso).
- **Situação física:** informação alfanumérica que identifica a situação da pista quanto à atividade atual. Os possíveis valores são oriundos da classe Situação Física, podendo ser: desconhecida (situação física desconhecida); abandonada (quando não há investimentos para sua recuperação ou manutenção); destruída (recuperação economicamente inviável, não sendo possível ser recuperada por ter sua estrutura fundamental comprometida); construída; em construção; planejada; construída (mas em obras); ou não aplicável.
- **Largura:** informação do tipo real que indica a largura oficial (ROTAER) da pista ou do ponto de pouso (em metros).
- **Extensão:** informação do tipo real que indica o comprimento oficial (ROTAER) da pista (em metros).
- **Altitude:** informação do tipo real que indica a altitude oficial (ROTAER) do ponto de referência do complexo (em metros).

As estruturas de dados disponibilizadas pela INDE são um conjunto mínimo de informações esperado para os elementos geográficos. Assim, é possível incluir novas informações de forma a expandir tal conjunto, já que a INDE não restringe a agregação de informações por parte de instituições oficiais.

Para apoiar essa ação e também a divulgação e a disponibilização de informações geoespaciais, é utilizado o Visualizador da Infraestrutura Nacional de Dados Espaciais (VINDE). Essa ferramenta visa facilitar o acesso às informações geoespaciais de forma agregada. Os dados

disponibilizados por ela podem ter diferentes formatos e a localização dos serviços pode pertencer a diferentes organizações.

Além da disponibilização de dados, a ferramenta permite a criação de mapas e a execução de diferentes funções, como visualização, navegação, consultas básicas e medição de distâncias nas superfícies. O projeto VINDE possibilita, portanto, que as instituições produtoras de dados georreferenciados os disponibilizem na INDE para toda a sociedade.

O VINDE foi desenvolvido pela Diretoria de Informática do IBGE. Em 2012, a iniciativa foi contemplada com o Prêmio TI & Governo, que é um concurso nacional que objetiva distinguir as melhores iniciativas de governo eletrônico no Brasil, abrangendo instituições federais, estaduais e municipais. Até 2017, foram disponibilizados dados espaciais de 13 instituições públicas em 37 temas.

Tabela 1 – Resumo dos dados identificados.

Fonte	Dado	Formato de disponibilização	Periodicidade de atualização	Existe documentação
ANAC	Cadastro de aeródromo	XLS	Mensal	Sim
	Horário de transporte	CSV, HTML, XLS	Diário	Sim
	Movimentação aeroportuária	CSV	Mensal	Sim
	Voo regular ativo	CSV	Mensal	Sim
	Tarifas aéreas	CSV	Mensal	Sim
	Estatísticas	XLS	Mensal	Não
	Aeródromo	WebService	Diário	Sim
DECEA	Cartas NOTAM	WebService	Diário	Sim
	Suplementos de AIP	WebService	Diário	Sim
	Nascer e pôr do sol	WebService	Diário	Sim
	METAR	WebService	Diário	Sim
	Desempenho Operacional	CSV	Trimestral	Não
SAC/MTPA	Comércio Exterior	CSV	Mensal	Sim
MIDIC	Indicadores de município	CSV	Annual	Não

Fonte: O autor (2018)

### 3 MÉTODO PARA CONSTRUÇÃO DA BASE DE DADOS

Construir bases de dados de maneira sistêmica e metódica, aplicando técnicas que possam ser reproduzidas, garante a evolução e a continuidade das bases. Nesse processo, o conhecimento sobre as informações que farão parte da base de dados é evidenciado, bem como suas restrições e limitações de uso.

A qualidade da base de dados desenvolvida (quanto ao fornecimento de informações) está ligada à qualidade e à disponibilidade das informações, sendo importante a definição de um processo para reger as atividades que circundam sua construção propriamente dita. Um processo documentado permite a reprodução das atividades mesmo em períodos de tempo diferentes, fazendo com que a construção de base de dados seja executada sempre da mesma forma.

Para suprir as demandas por dados que circundam projetos de planejamento do setor de transportes, neste trabalho, é proposto um método que auxilie a criação de bases de dados e a atualização de existentes. Tal método pode ser aplicado a qualquer área de transportes, desde que no estudo de planejamento exista a necessidade de construir-se uma base de dados ou mesmo de agregar novas informações em uma base de dados existente.

É recomendado que o método seja utilizado no início dos projetos de planejamento, principalmente após a identificação das primeiras de bases de dados necessárias para a sua execução. Nesse contexto, os dados necessários para o planejamento formarão uma base de dados que será construída de forma sistêmica, expansível e por meio de incrementos. Conforme apresentado na Figura 9, o método deve ser aplicado para cada dado identificado, e sua utilização permitirá a extração de informações consistentes para o projeto de planejamento.

Com a finalidade de apoiar os projetos para o planejamento do setor de transportes aéreo, construiu-se uma base de dados seguindo as atividades do método proposto. Tal base contempla informações das fontes de dados identificadas neste trabalho, com os dados dos períodos disponíveis, atendendo à fase de construção de base de dados, presente na elaboração de planos de sistemas aeroportuários e planos diretores. As atividades que compõem o método proposto são descritas a seguir:

1. **Identificar conjunto de dados:** esta atividade inicia-se com o surgimento da necessidade de uma base de dados com informações para o planejamento de um sistema de transporte. No contexto

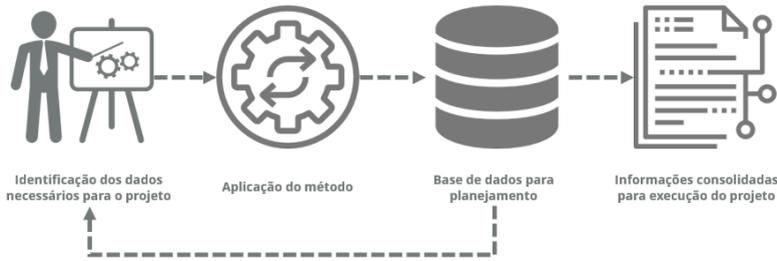


Figura 9 – Cenário de aplicação do método para construção de bases de dados.

Fonte: O autor (2018).

deste trabalho, como apresentado na seção 2.2, para a realização de planejamento do setor aéreo, são necessárias diversas informações; entre elas, pode-se citar localização do aeródromo, movimentação (passageiros, aeronaves, carga aérea e mala postal), horários de voos, rotas (origens e destinos), percentuais de atrasos e cancelamento de voos. Com base no conjunto de dados identificado, são evidenciadas as reais necessidades por informações, bem como os principais cenários de uso dos dados para o planejamento que será realizado. Ao reunir o conjunto de necessidades, a saber, os requisitos para o planejamento da base de dados, estes são evidenciados e registrados como forma de orientar o processo de construção da base. Durante essa atividade, a seguinte questão deve ser respondida: "Entre os dados disponíveis, quais são necessários à realização do planejamento de um sistema de transportes?"

2. **Identificar fontes de dados:** de posse das necessidades e do conjunto de dados para a construção da base, começa a busca e a identificação das fontes de dados. A busca deve ser realizada principalmente em *sites* e portais governamentais, pois são os veículos que disponibilizam informações oficiais. Além disso, é importante que as informações públicas estejam em formato aberto para permitirem utilização plena. No contexto do setor aeroportuário civil brasileiro, os principais portais que disponibilizam informações abertas são a ANAC e o IBGE; portanto, a busca por informações do setor pode ser iniciada por eles. É nessa atividade que se identificam as restrições de acesso, uso e divulgação dos dados (principalmente quando se tratam de dados

restritos ou confidenciais, pois em alguns casos, é possível obter autorização para uso dos dados). Como resultado, é esperado que as fontes de dados (que serão utilizadas para a construção da base de dados) sejam registradas em documento indicando conjunto de dados, nome da fonte e endereço de acesso. Durante essa atividade, a seguinte questão deve ser respondida: "Quais fontes de dados disponibilizam os dados necessários para a realização da coleta dos dados?"

3. **Coletar dados:** de posse das fontes, inicia-se o procedimento de coleta de dados e informações disponíveis. Independentemente da fonte de dados, cada conjunto de dados tem suas particularidades no acesso e na coleta. Nessa atividade, pode surgir a necessidade de estabelecimento de diversas alternativas para que as coletas de dados sejam efetivadas, pois é comum que, para cada fonte, exista um modo diferente de coleta, devido à maneira como cada instituição disponibiliza os dados. Normalmente, os dados estão disponíveis por meio de *webservices*<sup>1</sup> ou *download* de arquivos (por exemplo, planilhas ou arquivo de textos). Nessa atividade, é importante o registro de como cada dado é coletado, permitindo a repetição desse processo em caso de atualização dos dados. O registro deve ser realizado por meio da tabulação do nome do dado propriamente dito e do nome do documento que contém o dado ou *webservice* (URL), gerando um mapeamento dos dados com o respectivo local físico (diretório ou pasta no sistema de arquivos) em que se encontram. Tanto o nome do documento quanto o diretório em que está localizado devem ser padronizados, armazenados e registrados em um documento para facilitar a sua recuperação. Portanto, é importante que, na nomenclatura do arquivo ou *webservice*, seja fácil a identificação do conteúdo e de quando foi coletado. No início a coleta dos dados, deve ser criado um diretório para o armazenamento centralizado de todos os documentos e dados, pois espera-se que os dados necessários para a construção da base encontrem-se em um local físico único e estejam acessíveis para não impactar (por falta de acesso) as demais atividades da metodologia proposta. Durante essa atividade, deve ser respondida a seguinte questão: "Quais dados identificados e disponíveis foram coletados?"

---

<sup>1</sup> *Software on-line* que é acessado por meio de uma URL e que possui interface baseada em XML para descrever uma coleção de operações acessíveis através de rede. Os *webservices* são comumente utilizados no compartilhamento de dados entre *softwares* de plataformas e linguagens diferentes (GIRARDI, 2004).

4. **Analisar dados coletados quanto a estrutura:** quando os dados são desconhecidos ou novos, é fundamental a realização de uma análise quanto à sua estrutura. Essa atividade permite a identificação de campos fundamentais para a estruturação das informações, além da identificação de formas de agregação e relacionamentos entre as diversas fontes de dados. Espera-se que cada conjunto de dados seja documentado, identificando suas estruturas, restrições e os tipos de dados. No caso de uma atualização ou de manutenção de um conjunto de dados, é importante que os novos dados sejam confrontados com os existentes, permitindo a adoção de estratégias de reaproveitamento de estruturas e garantindo que os dados continuem sendo disponibilizados no formato previamente identificado e estabelecido. Na primeira vez em que essa atividade é executada, o resultado é o modelo de dados que deve compor a base; nas próximas vezes (atualização ou manutenção da base), o objetivo é garantir que os dados estejam compatíveis com o que foi definido anteriormente. Durante essa atividade, deve ser respondida a seguinte questão: "Todas as estruturas necessárias dos dados coletados são conhecidas e suficientes para o armazenamento dos dados a serem coletados?".
  
5. **Analisar dados coletados quanto ao conteúdo:** de posse dos dados, nesta atividade, eles são analisados quanto ao conteúdo, visando à identificação de padrões e restrições. Mesmo que comumente essa atividade seja realizada em paralelo à anterior (analisar dados coletados quanto a estrutura), é importante que seja executada em sequência, a fim de permitir a validação da estrutura de dados já estabelecida. Nesse momento, para que a padronização dos dados seja concretizada e para garantir a qualidade dos dados que devem compor a base, espera-se que os dados sejam mantidos de maneira uniforme, eliminando dados duplicados, incoerentes, incompatíveis e não necessários. Quando uma base de dados já existir, deve-se validar as referências e equivalências entre os dados existentes na base. O resultado é um conjunto de dados padronizados e organizados. Durante essa atividade, deve ser respondida à seguinte questão: "Todas as regras e restrições necessárias relacionadas aos dados coletados são conhecidas?".
  
6. **Criar ou atualizar estrutura de banco de dados:** de posse da estrutura dos dados e seus tipos, nesta atividade, é realizada a criação ou a atualização da base de dados propriamente dita. Ao executar essa atividade pela primeira vez, o serviço de banco

de dados é criado e devidamente configurado; são desenvolvidas também as estruturas de tabelas e os relacionamentos entre elas (campos criados nas tabelas que fazem referência a um valor de outra tabela). Quando novos dados são coletados ou o formato da estrutura identificada inicialmente é modificado pela fonte dos dados (são eliminadas ou adicionadas novas informações), existe a necessidade de criar uma estrutura de dados compatível para o armazenamento. Assim, as estruturas de tabelas são atualizadas ou são elaboradas novas tabelas para que se torne possível a inclusão desses novos dados na base existente. Como resultado dessa atividade, tem-se a base dados disponível e acessível, possibilitando centralizar todos os dados coletados. Durante essa atividade, deve ser respondida a seguinte questão: "A base de dados e as estruturas necessárias estão disponíveis para armazenamento e uso dos dados?"

7. **Registrar metadados:** com a base de dados criada, existe a necessidade de registrar quais são as fontes e os dados existentes na base de dados. Para isso, sugere-se que as informações de metadados sejam tabuladas (na própria base de dados ou fora dela) para registro e armazenamento das informações acerca da base de dados. Nesse registro, devem constar informações sobre as tabelas e colunas, e seu acesso deve estar disponível sempre que necessário para garantir que os usuários da base consigam entender e utilizar os dados da melhor forma. Sempre que houver atualização na estrutura da base de dados, o conjunto de informações que compõe os metadados deve ser atualizado e disponibilizado aos usuários. Como resultado dessa atividade, espera-se um conjunto de informações descritivas a respeito de cada tabela que compõe a base de dados. No âmbito dos dados geoespaciais, é importante que seus metadados estejam aderentes ao proposto pela INDE. Nesse sentido, é importante a realização de consulta ao documento "Perfil de Metadados Geoespaciais do Brasil", buscando as proposições de metadados para os elementos geoespaciais a serem inseridos na base de dados. Durante essa atividade, deve ser respondida a seguinte questão: "Todas as tabelas e colunas estão documentadas?"
8. **Inserir ou atualizar dados:** para esta atividade, é realizado um procedimento de transformação (converter o formato disponível para um modo compatível com a estrutura da base de dados) e carga dos dados para que estes fiquem compatíveis com a estru-

tura já desenvolvida. Quando essa atividade é executada pela primeira vez, é feito um processo de inserção de forma direta, ou seja, os dados são simplesmente inseridos na base de dados. Na situação de já existirem dados na estrutura de tabelas do banco, tais dados devem ser substituídos. Nesse contexto, é possível evidenciar três situações:

- **Dado coletado anula todos os dados existentes:** primeiramente, elimina-se os dados existentes, deixando as tabelas sem nenhum dado; em seguida, é realizado o carregamento dos novos dados.
- **Dado coletado é complemento do conjunto de dados existente:** a inserção ocorre de forma direta, porém, deve existir cuidado com os dados existentes, pois estes juntamente com os novos formarão o conjunto de dados final da base de dados.
- **Dado coletado é uma atualização em registros específicos:** nesta situação, é necessário garantir que somente os registros corretos serão atualizados. Para isso, deve ser feita uma comparação entre os novos dados e os existentes para que a atualização da base seja executada somente nos registros corretos. Ainda, deve-se comparar as colunas-chave, que garantem a unicidade do registro.

O procedimento de inserção ou atualização dos dados pode ser executado de forma manual ou automatizada. De modo manual, são geradas tabelas (com o mesmo formato do dado coletado) na base de dados com caráter temporário e realizado o carregamento dos dados nesse conjunto de tabelas. Em seguida, é feita a transformação e, por fim, o carregamento dos dados já nas tabelas finais, as quais irão armazená-los na base. Já no procedimento automatizado, são desenvolvidas rotinas de *software* com a função de carregar, transformar e transferir os dados para as tabelas da base, seja para inserção inicial ou atualização. Para cada conjunto de dados, pode haver uma forma diferente, seja manual ou automatizada, que facilite essa demanda. Mesmo que essa atividade não seja restrita a um modo de execução, recomenda-se o desenvolvimento de rotinas automatizadas, pois, desse modo, garante-se que a atividade sempre será executada da mesma forma, não havendo interferências subjetivas. Como resultado dessa atividade, espera-se que o banco de dados definido esteja com todas as informações coletadas atualizadas. Durante essa atividade, deve

ser respondida a seguinte questão: "Todos os dados contidos na base de dados estão atualizados?"

9. **Validar dados:** é fundamental a execução de uma atividade para garantir que a base de dados criada esteja compatível com os dados disponibilizados pela fonte de dados. Portanto, nesta atividade, deve ser realizada uma conferência dos dados, confrontando os registros existentes na base de dados com os dados coletados. Para tanto, devem ser estabelecidos alguns indicadores que permitam a comparação com os dados coletados, por exemplo, quantidade de registros (da tabela em relação ao dado coletado) ou soma de colunas numéricas (da tabela em relação à soma da coluna equivalente no arquivo). Assim como a atividade "Inserir ou atualizar dados", esta também pode ser realizada de forma manual ou automatizada. De modo manual, existe a necessidade da execução de um conjunto de consultas à base de dados, para que seus resultados sejam confrontados com os resultados do arquivo ou dos *webservices* que possuem os dados coletados. De forma automatizada, é possível incluir, na rotina de inserção ou atualização de dados, um relatório comparativo dos dados, o qual pode ser gerado durante o processo de carga dos dados. Tais relatórios indicam o quão fiel a base de dados está em relação ao dado coletado, evidenciando a autenticidade dos dados contidos na base. Durante essa atividade, deve ser respondida a seguinte questão: "Os dados carregados na base de dados estão de acordo com os coletados?"
10. **Desenvolver sistema para uso dos dados:** com a base de dados devidamente validada, é importante que ela seja disponibilizada aos usuários finais. A atividade de disponibilização de um sistema faz com que se estabeleçam estratégias para acesso aos dados pelos usuários finais. Existem inúmeras formas para a disponibilização da base, tais como construção de *webservices*, acesso direto por meio de planilhas eletrônicas e até mesmo acesso aos dados diretamente na base – essa última alternativa para usuários com conhecimento em linguagens de banco de dados. No entanto, comumente, as bases de dados são disponibilizadas por meio de um *software* específico; espera-se, nesse caso, a oferta de um sistema computacional que ofereça acesso aos dados. O mais importante nessa atividade é que os dados estejam acessíveis por meio de um sistema, para que os usuários possam realizar suas atividades de planejamento. Durante essa atividade, deve ser respondida

a seguinte questão: "Um sistema computacional para acesso aos dados está disponível?".

Para a proposição do método, utilizou-se como base os conceitos apresentados no processo ETL, *Data Warehouse* e *Data Mining* (como apresentado respectivamente nas seções 2.1.1, 2.1.2 e 2.1.4).

As etapas do processo ETL relacionadas à extração dos dados estão distribuídas nas atividades "Identificar conjunto de dados", "Identificar fontes de dados" e "Coletar dados". As etapas do processo ETL relacionadas à transformação dos dados estão distribuídas nas atividades "Analisar dados coletados quanto a estrutura", "Analisar dados coletados quanto a conteúdo" e "Criar ou atualizar estrutura de banco de dados", pois tratam da análise dos dados para viabilizar sua formatação e transformação em modelo próprio. As etapas do processo ETL relacionadas à carga dos dados estão distribuídas nas atividades "Inserir ou atualizar dados", "Registrar metadados" e "Validar dados", pois garantem que os dados sejam carregados na base. Na atividade "Desenvolver sistema para uso dos dados", são desenvolvidas visões específicas dos dados (*data warehouse* ou *data mart*), bem como a exploração de novas informações e conhecimentos (*data mining*).

O alinhamento entre as atividades do método e a literatura podem ser observados na Figura 10, que apresenta o fluxo de atividades do método proposto, evidenciando as etapas que estão relacionadas a cada item da literatura.

O método proposto pode ser aplicado no desenvolvimento de uma base de dados nova, na inclusão de novos dados em uma base de dados existente e ainda na atualização de dados e estrutura de uma base. Em todos os modos de utilização, foram identificadas ações que podem ser automatizadas por meio de uma rotina de *software*, como é caso das atividades "Coletar dados", "Analisar dados coletados quanto a estrutura", "Analisar dados coletados quanto ao conteúdo", "Criar ou atualizar estrutura de banco de dados", "Registrar metadados", "Inserir ou atualizar dados", "Validar dados" e "Desenvolver sistema para uso dos dados".

No entanto, é importante salientar que a primeira execução das atividades do método deve ser feita de forma manual, isso porque o desenvolvimento de um *software* ou de uma rotina que execute as atividades de forma automatizada exige, antes, a identificação de *layouts* de dados, tipos de dados e regras de uso de dados. Tais elementos são como para orientar a elaboração de funcionalidades, permitindo o desenvolvimento de um *software* que saiba interpretar e analisar o conjunto de dados. De forma geral, permitem que o sistema tenha

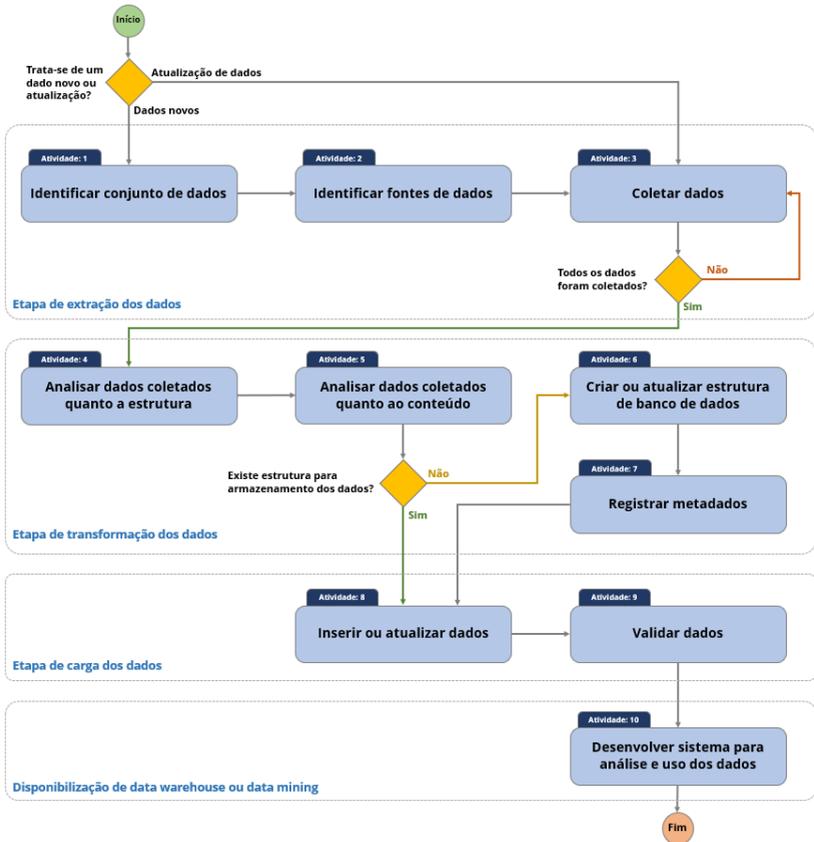


Figura 10 – Fluxo de atividades do método para construção de bases de dados.

Fonte: O autor (2018).

informações necessárias para que possa tratar/manipular os dados da melhor forma. Portanto, primeiramente, executam-se as atividades de forma manual e, posteriormente, inicia-se a automatização delas.

Quando o método for aplicado para a inclusão de novos dados em uma base existente, a atividade denominada "Criar ou atualizar estrutura de banco de dados" tem sua execução realizada de forma parcial, isso porque não será necessário criar e configurar os serviços de banco de dados (utiliza-se os existentes). Assim, existe apenas a necessidade de desenvolvimento ou de evolução das novas estruturas e tabelas para

o armazenamento dos novos dados, bem como para o relacionando entre as novas tabelas e aquelas existentes. O mesmo ocorre no caso de atualização de dados e estrutura de uma base, ou seja, nesse caso, a atividade "Criar ou atualizar estrutura de banco de dados" também é realizada apenas parcialmente. Assim, aplicam-se as atividades relacionadas à coleta, à atualização e à validação dos dados

Independentemente do cenário de uso do método, as atividades "Coletar dados", "Analisar dados coletados quanto a estrutura", "Inserir ou atualizar dados" e "Validar dados" são periódicas, ou seja, sempre que houver necessidade de atualização dos dados, elas devem ser executadas. Tal demanda surge quando se necessita de um dado atualizado para um planejamento específico ou pela identificação de disponibilidade um dado mais atualizado. Como essa necessidade pode surgir a qualquer momento, é recomendado que as atividades supracitadas sejam automatizadas, permitindo que possam ser repetidas inclusive por indivíduos com pouco conhecimento acerca de bases de dados.

## 4 APLICAÇÃO DO MÉTODO PARA CONSTRUÇÃO DE BASE DE DADOS

Para a construção da base de dados para apoio a projetos e iniciativas de planejamento do setor de aviação civil, foi estabelecido como guia o método proposto neste trabalho. Assim, inicialmente, foram identificados os dados necessários; em seguida, foram explorados os portais governamentais que fornecem dados relevantes para o contexto deste trabalho. O conjunto de dados necessários foi coletado e organizado em um sistema de arquivos para posterior execução das outras atividades do método. Após a reunião dos dados, foi iniciada a análise do conjunto, a fim de determinar o padrão e o formato das tabelas que armazenariam os dados. Foi realizada então a inserção dos dados na base e a disponibilização desta para uso.

Para a aplicação do método, foram utilizados: o Microsoft Excel, para o tratamento dos dados; o Enterprise Architect, para a modelagem das estruturas de tabelas que iriam compor o banco de dados; o SQL Manager for SQL Server, para utilização do banco de dados (geração de tabelas e inserção dos dados). A modelagem das estruturas de tabelas foi desenvolvida utilizando-se a notação chamada "Pés de galinha", pois esta ilustra a relação entre as entidades de um diagrama de banco de dados por meio símbolos em ambas as extremidades, os quais descrevem a cardinalidade da relação entre as entidades.

### 4.1 IDENTIFICAÇÃO CONJUNTO DE DADOS E FONTES DE DADOS

A identificação das fontes de dados para este trabalho foi realizada por meio de buscas em *sites* e portais governamentais que disponibilizam informações que se adequam às políticas de dados abertos. A busca por fontes de dados foi guiada pelo objetivo geral deste trabalho, qual seja, a proposição de uma base de dados para apoiar o planejamento do setor de aviação civil. Na tabela 2, é possível visualizar a lista de dados e fontes primárias de informações encontradas.

A prioridade deste trabalho era a obtenção de dados que seguissem as regras de dados abertos quanto a formato de disponibilização, meio de acesso e documentação. Porém, algumas fontes encontradas que não se adequavam a tais regras foram analisadas a fim de que se pudesse avaliar sua aderência com o trabalho.

Tabela 2 – Registro de fontes primárias de dados.

Conjunto de dados	Fonte	Endereço
Indicadores de municípios	IBGE	ftp://ftp.ibge.gov.br
Cadastro de aeródromos	ANAC	http://www.anac.gov.br
Voo regular ativo	ANAC	http://www.anac.gov.br
Dados estatísticos	ANAC	http://www.anac.gov.br
Horário de transporte	ANAC	http://www.anac.gov.br
Tarifas aéreas	ANAC	http://www.anac.gov.br
Estatísticas	INFRAERO	http://www.infraero.gov.br
Desempenho Operacional	SAC/MTPA	https://horus.labtrans.ufsc.br/
NOTRAN	DECEA	http://dados.gov.br

Fonte: O autor (2018).

## 4.2 COLETA DOS DADOS

A atividade de coleta das informações foi realizada assim que finalizada a busca por fonte de dados. Com a finalidade de estruturar a coleta dos dados, bem como executar tal atividade de forma sistêmica, foram gerados dois elementos de apoio: uma planilha de controle da coleta e um diretório (pasta no sistema de arquivo) para armazenamento centralizado de todos os dados. Além disso, foi aplicada uma nomenclatura para a organização dos dados coletados por data de coleta, fonte e nome. Tal nomenclatura fez com que os arquivos tivessem o seguinte formato "**aaaa**mm**dd**\_FONTE\_NomeDocumento", onde:

- **aaaa**: ano em que o dado foi coletado;
- **mm**: mês em que o dado foi coletado;
- **dd**: dia em que o dado foi coletado;
- **FONTE**: sigla do órgão fonte do dado coletado; e
- **NomeDocumento**: nome do documento sem a utilização de caracteres especiais, sem espaçamento e sempre com a primeira letra das palavras que compõem o nome em maiúscula.

Cada fonte de dado primário requisitou a coleta de dados de uma forma fonte, pois cada *site* ou portal disponibiliza a informação de um modo distinto (esse seria um desalinhamento à regra de dados abertos). Durante a coleta, foram identificados manuais, relatórios e documentações que registram detalhamentos sobre o uso dos dados.

Esse conjunto de documentos foi utilizado como subsídio para a atividade de análise da estrutura e do conteúdo de cada dado e permitiu compreender e entender regras acerca dos dados coletados. Nesta seção do trabalho, são apresentadas as formas de coleta dos dados aplicadas para a construção base de dados de apoio ao planejamento aeroportuário brasileiro. Na tabela 3, é possível visualizar os registros dos dados coletados durante a atividade.

O portal do IBGE disponibiliza informações sociais e econômicas acerca dos municípios brasileiros – inclusive, é possível encontrar uma mesma informação de formas e em locais diferentes no portal. O Instituto disponibiliza um FTP ([ftp://ftp.ibge.gov.br//Pib\\_Municipios](ftp://ftp.ibge.gov.br//Pib_Municipios)) em que estão armazenadas todas as pesquisas sobre o PIB de municípios brasileiros em formato de planilha (CSV). Além dos dados propriamente ditos sobre os municípios, nesse FTP, estão disponíveis notas metodológicas da coleta.

Já o conjunto de dados disponibilizado pela ANAC encontra-se no portal da agência de forma centralizada, isso por meio de uma página que fornece os *links* para páginas específicas de cada dado. As planilhas de dados cadastrais de aeródromos públicos e privados estão disponíveis no mesmo local, porém, cada conjunto de aeródromos encontra-se em uma planilha diferente.

Os dados de VRA estão disponíveis em formato mensal, sendo os de janeiro de 2000 até dezembro de 2015 disponíveis no formato CSV em um arquivo compactado; já a partir de janeiro de 2016, os dados estão no formato CSV sem compactação. Os dados estatísticos estão disponíveis a partir de janeiro de 2000, sendo disponibilizados de duas formas: arquivos anuais (CSV) e base completa (ZIP) com um arquivo CSV dentro. Para os dados estatísticos, utilizou-se a base completa, pois garante a atualização dos dados de todos os períodos.

Os dados de HOTRAN podem ser coletados no portal da ANAC de duas formas: acesso ao arquivo do dia corrente na planilha de voos autorizados vigentes (planilha 1.5); e por um sistema *on-line* que permite a visualização dos históricos. Como o interesse nesse conjunto de dados concentra-se nas séries históricas, os dados foram coletados do sistema *on-line* de HOTRAN disponibilizado pela ANAC.

Todos os dados coletados foram armazenados no diretório; além disso, como proposto no método, foi realizada a tabulação desses dados coletados de forma individual, apoiando a análise e a posterior construção da base de dados. Tal atividade mostrou-se relevante pois deu subsídios para a homologação e o registro das informações que farão parte da base de dados.

### 4.2.1 Análise dos dados coletados

Como proposto no método de criação de base de dados para o planejamento aeroportuário brasileiro, as atividades "Analisar dados coletados quanto a estrutura" e "Analisar dados coletados quanto ao conteúdo" são realizadas de forma sequencial. Portanto, a explanação de como foram executadas é apresentada na mesma seção. Definiu-se esse modo de organização do conteúdo devido ao grande volume de dados, buscando a melhor compreensão sobre o registro dessas atividades. Para todos os dados coletados, foi realizada uma análise quanto à sua estrutura e ao seu conteúdo, buscando estabelecer um formato adequado para seu armazenamento na base de dados, bem como identificar regras de utilização de tais dados.

### 4.2.2 Dados de estados e municípios

No conjunto de dados disponíveis para municípios brasileiros, os indicadores aparecem de forma replicada, pois cada linha da planilha representa um registro por ano e município. Nesse contexto, existem informações cadastrais sobre outras entidades geopolíticas, tais como unidade federativa, mesorregião e microrregião, sendo todas identificadas por nome da entidade e seu código. Na tabela 4, é possível visualizar o resultado da análise da estrutura destes dados.

A coleta, a compilação e a organização dos dados do IBGE foi realizada na seguinte ordem: unidade federativa, mesorregião, microrregião, município e indicadores anuais por município. Esses dados foram inicialmente manipulados utilizando-se a ferramenta Microsoft Excel; assim, para cada conjunto de dados, foi gerada uma planilha específica, para que posteriormente pudessem ser utilizados na construção da base de dados. A organização dos dados por tipo de entidade geopolítica, separadamente, permitiu identificar as regras e os formatos dos dados para uma possível automatização desse processo.

Na planilha do IBGE, não estão disponíveis as siglas de unidades da federação. Como comumente tal informação é utilizada nos conjuntos de dados externos ao Instituto, definiu-se que seria relevante a existência das siglas. Para a coleta dessa informação, realizou-se uma consulta no portal IBGE (2017a), e os resultados (siglas) foram inseridos na planilha gerada para a unidade da federação. Na tabela 5, é possível visualizar como foi o resultado dessa compilação de unidades da federação.

Para a estrutura de tabela de unidades da federação, foi estabelecido que o código disponibilizado pelo IBGE (código único e numérico) passaria a ser o identificador da tabela. Assim, todas as tabelas da base de dados que forem referenciar uma unidade da federação passariam a utilizar tal código. Os campos sigla e nome são textuais e, como não existem unidades da federação com nomes iguais, para eles, foi estabelecido que conteriam somente valores únicos.

Para o procedimento de compilação dos dados de municípios propriamente ditos, os dados da planilha do IBGE foram transferidos para uma nova planilha, os duplicados foram removidos e foram mantidas somente as colunas que faziam referência ao município. Portanto, foram mantidos na nova planilha o código do município, o nome e o código da unidade da federação da qual o município faz parte. Na tabela 6, é possível visualizar alguns registros do resultado dessa compilação.

Para a criação da tabela de municípios, foi estabelecido que o código disponibilizado pelo IBGE (código único e numérico) seria o identificador dos registros existentes na tabela. O campo nome representa o nome do próprio município; já o código da unidade da federação representa um valor de referência aos registros da tabela de unidade da federação, sendo uma forma de ligar o município à unidade da federação da qual ele faz parte.

Para os dados de indicadores de municípios, não houve a necessidade de tratamento quanto a valores duplicados devido ao fato de a estrutura, na planilha do IBGE, já encontrar-se no formato determinado para tais dados, sendo eles estruturados em anual e municípios. Sendo assim, para esse conjunto de dados, foram removidas as colunas que não fazem parte do contexto de indicadores por municípios. Na figura 7, é possível visualizar alguns dos registros resultantes da compilação de indicadores dos municípios.

Visto a quantidade de indicadores disponíveis para cada município (total de nove), além da possibilidade de existência de outros novos indicadores, optou-se por desenvolver uma estrutura de tabelas que permitisse a escalabilidade. Sendo assim, foi gerada uma estrutura para o armazenamento dos indicadores e outra que armazena os registros de um valor para cada indicador, ano e município.

Na estrutura de tabelas para indicadores, foram geradas colunas para código de identificação, nome (nome abreviado) e descrição (um descritivo do que é o indicador). Na tabela associativa (faz a relação entre municípios e indicadores), foram criadas colunas de identificador do município, identificador do indicador, ano de referência e valor (valor que o município registrou para o ano e indicador).

Com o conjunto de tabelas apresentado, foi possível estabelecer uma estrutura para a base de dados acerca das informações das entidades geopolíticas que fazem parte do contexto deste trabalho. Na figura 11, é possível visualizar um modelo estrutural das tabelas e de seus relacionamentos, que farão parte da base de dados.

Mesmo que o método proposto não imponha restrições quanto ao modelo a ser adotado para a criação de continuidade da base de dados, é importante a automatização do processo para que este seja reproduzido por qualquer indivíduo. Para os dados de municípios do IBGE, após a execução das atividades de forma manual, foi possível identificar regras e padrões para que o processo de atualização e manutenção desses dados seja automatizado. Nesse sentido, o único requisito para esses dados é que já estejam coletados; após isso, é possível realizar todo o processo de carga por meio de um *software* seguindo os seguintes passos:

- carregar arquivo com indicadores;
- analisar se existem as colunas referentes aos indicadores;
- verificar se o arquivo seria uma atualização dos dados existente ou complemento;
- inserir os registros novos ou atualizar os existentes.

#### 4.2.3 Dados de cadastro de aeródromos

Nas planilhas de aeródromos públicos e privados, estão contidas as informações de localização, operação e infraestrutura; tais dados foram consolidados em uma única planilha. Nesse processo, identificou-se que em ambas planilhas a coluna tipo de aeródromo sempre continha o valor "aeroporto"; desse modo, foi realizada uma alteração, na qual os dados de aeródromos públicos passaram a receber o valor "Nacional Público" e os privados passaram a receber "Nacional Privado". Na tabela 8, é possível visualizar o resultado da análise da estrutura desses dados.

Após consolidar os dados de aeródromos (unir a planilha de aeródromos públicos com a de aeródromos privados), identificou-se que muitos dados deveriam ser subdivididos e organizados em estruturas separadas por grupo de informações comuns. Desse modo, foi proposta uma tabela contendo informações de ICAO, nome do aeródromo, município (código do município), tipo de aeródromo (código do tipo de

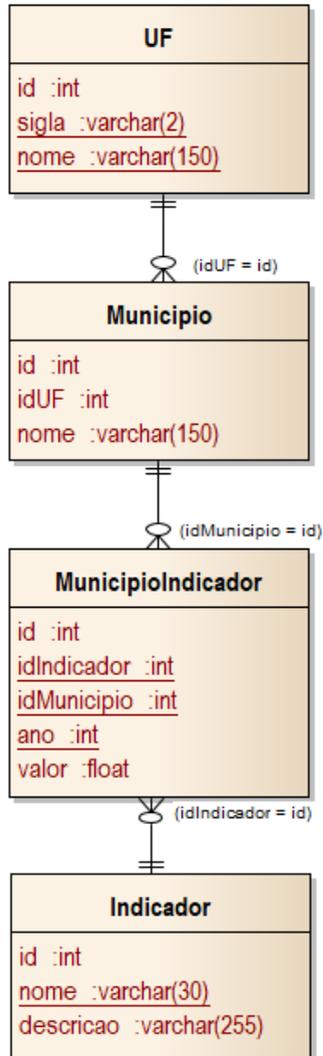


Figura 11 – Modelo estrutural do banco de dados para municípios.

Fonte: O autor (2018).

aeródromo), latitude, longitude e altitude. Além desses campos, foi estabelecido um identificador (código numérico) único, que seria a chave primária da tabela.

Durante o processo de análise dos dados de aeródromos, foi verificado que muitos municípios (município de atendimento do aeródromo) não estavam com a mesma nomenclatura disponível no IBGE. Por isso, para cada aeródromo, não houve relação direta entre o nome do município e a unidade da federação com os disponíveis pelo IBGE. A principal causa do problema é a forma como o nome do município estava escrito na base da ANAC; assim, foi necessária a correção da grafia dos nomes dos municípios. As coordenadas geográficas da planilha da ANAC estavam com valores em graus e minutos decimais e foram convertidas para graus decimais; para tanto, foi utilizado o *software* IBGE TCGeo. Na tabela 9, é possível visualizar alguns dos registros resultantes da compilação dos dados cadastrais de aeródromos.

As regras de operação dos aeródromos públicos e privados são iguais na prática, porém, as planilhas apresentam termos diferentes para um mesmo elemento. A partir desse cenário, foi desenvolvida uma padronização desses dados, e se passou a ter uma leitura única da informação. Para a parte de regras de operação, foi designada uma estrutura que armazena as regras existentes (IFR diurno, IFR noturno, VFR diurno e VFR noturno); nela, criou-se um campo para o identificador e outro para a descrição.

Durante a análise dos dados de estrutura dos aeródromos, foi identificado que eles podem conter mais de uma pista de pouso e decolagem. Para que não fosse fixada na estrutura o número de pistas dos aeródromos, foi elaborada uma tabela que relaciona os dados de pistas com o aeródromo. Além disso, identificou-se que as pistas podem ter tipos de pavimentos diferentes, então, criou-se uma tabela para o armazenamento dos tipos de pavimentos existentes. Assim, as estruturas de armazenamento de pista contêm código identificador do aeródromo, número da pista, designador (código da cabeceira A e código da cabeceira B), comprimento, largura, resistência e pavimento. Na tabela 10, é possível visualizar alguns dos registros resultantes da compilação dos dados de pista.

Com o conjunto de tabelas apresentados para aeródromos, foi possível manter na base de dados as informações cadastrais dos aeródromos públicos e privados da ANAC. Na figura 12, é possível visualizar o relacionamento entre as tabelas desse contexto da base de dados.

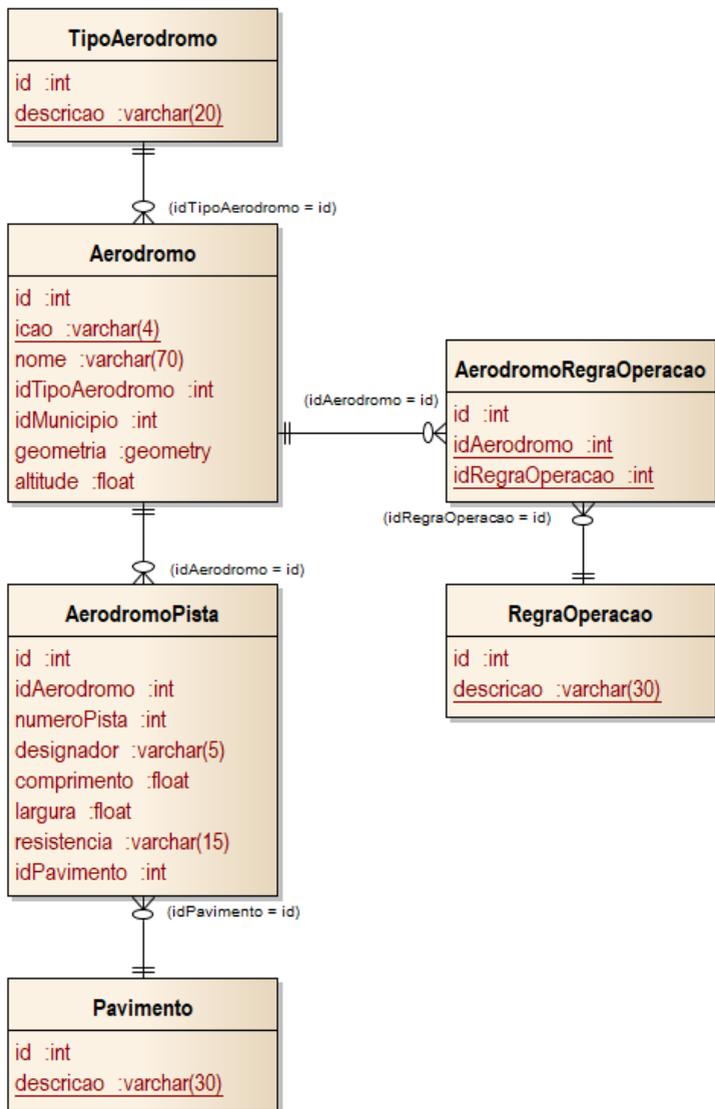


Figura 12 – Modelo estrutural do banco de dados para aeródromos.  
Fonte: O autor (2018).

#### 4.2.4 Dados de horário de transporte

Para os dados de HOTRAN, o processo de análise teve início com a avaliação da estrutura dos arquivos diários baixados (um total de 2.310 arquivos). Para isso, foi desenvolvido um *script* utilizando a linguagem de programação Java, com a finalidade de varrer vários arquivos com extensão CSV e indicar se eles possuíam a mesma estrutura de cabeçalhos. Após a confirmação de que todos os arquivos eram iguais, utilizou-se um deles como referência para a avaliação dos tipos de dados. Na tabela 11, é possível visualizar o tipo de dado para cada coluna da planilha de dados disponibilizada.

Durante a análise dos dados de companhia aéreas, verificou-se que no HOTRAN existiam apenas companhias que contemplavam os dados analisados na base estatística. Portanto, tais dados foram padronizados para que fosse possível alimentar a estrutura proposta de companhia aérea. Durante essa avaliação, foram considerados sigla e nome da companhia, pois são os elementos que permitem a identificação de registros únicos.

Para natureza da operação, a movimentação trata os dados somente como doméstico ou internacional. No HOTRAN, existem novas categorias de natureza da operação, que são Cargueiro Doméstico, Cargueiro Internacional, Linha Aérea Sistemica e Rede Postal. Esses registros foram organizados para a alimentação da estrutura de dados de natureza da operação.

Utilizando a ferramenta Excel, foram agrupados os dados de aeródromos (origem e destino) existentes no HOTRAN em uma coluna nova na planilha, para que se pudesse remover os valores duplicados. O resultado desses aeródromos foi confrontado com os registros existentes para a estrutura de cadastro de aeródromo, sendo que nenhum novo aeródromo foi encontrado. Do mesmo modo que na base estatística, foram mantidos os registros de aeródromos nacionais e internacionais que não estavam presentes nas listagens da ANAC, para que fosse possível conservar a relação de voos autorizados vigentes com o disponibilizado pela ANAC.

A estrutura de HOTRAN proposta para o armazenamento de dados no banco é similar àquela disponibilizada pela ANAC. Os dados de companhia aérea, aeródromos (origem e destino) e natureza de operação são elementos que possuem correlação com outras estruturas; desse modo, tais dados fazem referência a tais tabelas. Na figura 13, é possível visualizar a relação entre as estruturas propostas para os dados de HOTRAN.

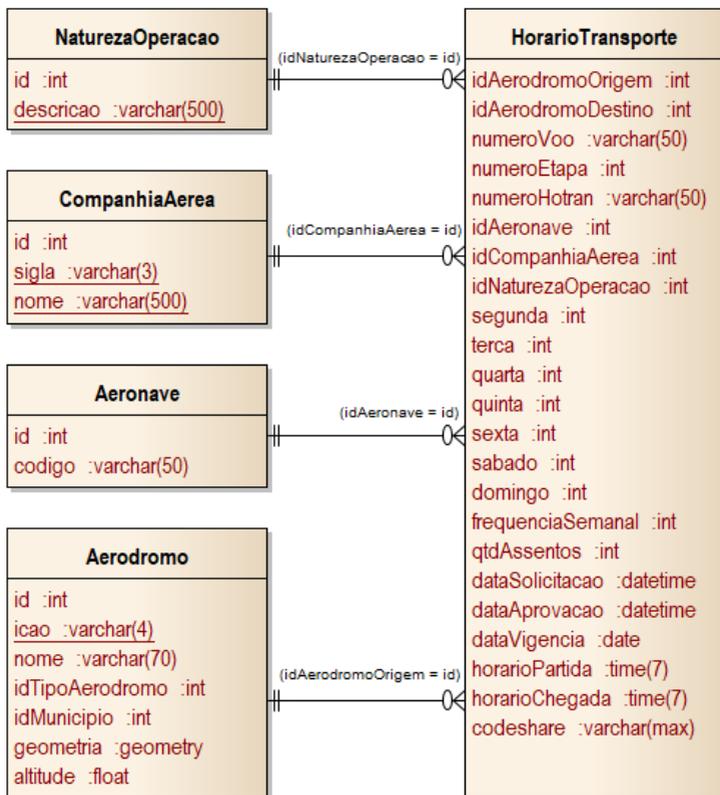


Figura 13 – Modelo estrutural do banco de dados para Horário de transporte.

Fonte: O autor (2018).

Verificou-se durante a análise dos arquivos de HOTRAN que, para apoiar o processo de inserção e atualização dos dados, era importante que a ferramenta desenvolvida para coletar os dados realizasse o armazenamento direto no banco de dados, pois eram muitos arquivos e seria complexa a sua homologação.

#### 4.2.5 Dados de movimentação aeroportuária

Para os dados estatísticos da ANAC, o processo de análise e padronização foi realizado pela base consolidada, como citado na atividade de coleta de dados. A estrutura desses dados conta com informações de movimentação de passageiros, cargas, aeronaves e correio (ou mala postal). Primeiramente, foram identificados os tipos de dados para cada coluna da planilha disponibilizada, como pode ser visualizado na tabela 12.

A primeira análise realizada nos dados de movimentação foi quanto aos aeródromos existentes. Identificou-se que existiam muitos aeródromos que não faziam parte do contexto deste trabalho, a saber, os internacionais. No entanto, para chegar-se a uma visão completa de sistema aeroportuário, dados de aeródromos que não estão na listagem de públicos e privados – e cuja maioria possui movimentação relacionada aos aeródromos das listagens pública e privada da ANAC – foram tratados para futura inserção na base dados.

Após identificar os aeródromos que não estão nas listagens da ANAC, tentou-se ao menos identificar se eles estavam localizados no Brasil. Para isso, foram consultados os aeródromos cujo código ICAO tinha as iniciais SB, SD, SI, SJ, SN, SS, SW (ICAO, 2006); o resultado dessa busca foram aeródromos nacionais. Desse modo, foram identificadas duas novas categorias para tipo de aeródromo: Nacional sem cadastro e Internacional. Ainda na análise de código ICAO na base de movimentação da ANAC, foram identificados códigos ICAO vazios, 0 e 0000; a partir disso, na padronização dos dados, os registros de movimentação com tais códigos ICAO inválidos (na origem ou no destino) foram descartados.

No conjunto de dados de movimentação, identificou-se que na coluna Natureza da operação existiam registros que se repetiam, no caso, Doméstica e Internacional. Assim, criou-se uma estrutura para armazenamento de natureza da operação.

O grupo de voo permite classificar a finalidade das rotas aéreas. Para movimentação, esses dados podem assumir os valores Improdutivo, Não se aplica, Não regular e Regular. Para seu armazenamento na base de dados, definiu-se uma estrutura com um campo de identificador e outro de descrição (recebe os grupos de voos possíveis).

As companhias aéreas são responsáveis pela execução do voo propriamente dito; no contexto da movimentação, existem 197 companhias aéreas que operam voos no Brasil. Para seu armazenamento na base de dados, foi desenvolvida uma estrutura com um campo de identi-

gador, um para sigla da companhia e outro para nome. Na tabela 13, é possível visualizar alguns dos registros resultantes da compilação dos dados de companhias aéreas.

Para o armazenamento da movimentação propriamente dita, foi elaborada uma estrutura de tabela que possui vínculos com companhia aérea, natureza da operação, grupo de voo e aeródromo (um para origem e outro para destino). Além dos vínculos associativos, a tabela também disponibiliza campos para indicação de período dos dados, indicadores da movimentação e uma data de referência dos dados.

Durante a análise dos dados, verificou-se que, para esse conjunto de dados, o processo de importação poderia ser realizado diretamente entre as tabelas estruturadas e as temporárias, pois a relação entre os dados dá-se de forma direta com as estruturas auxiliares. Além disso, com o conjunto de tabelas apresentado para movimentação, foi possível estabelecer uma estrutura interligada para manter na base de dados as informações acerca da movimentação. Na figura 14, é possível visualizar o relacionamento entre as tabelas desse contexto da base de dados.

A partir dessa estrutura para movimentação, durante o processo de análise dos dados, identificou-se que alguns registros seriam descartados automaticamente devido à falta de valor nos campos para os quais foi definido que não seriam aceitos valores nulos, tais como companhia aérea, aeródromo de origem, aeródromo de destino, natureza da operação, grupo de voo, ano e mês. Com esse processo, a base completa passou de 754.935 registros para 754.928 registros padronizados.

#### 4.2.6 Dados de voo regular ativo

Para os dados de VRA (voo regular ativo), foi adotado um processo de análise de estrutura similar ao do HOTRAN, no qual se iniciou avaliando a estrutura dos arquivos mensais baixados (um total de 182 arquivos). Para isso, o *script* utilizado para os dados de HOTRAN foi usado para varrer vários arquivos com extensão CSV e indicar se eles possuíam a mesma estrutura de cabeçalhos. Após a confirmação de que todos os arquivos eram iguais, utilizou-se um deles como referência para a avaliação dos tipos de dados. Na tabela 14, é possível visualizar os tipos de dados para cada coluna da planilha de dados disponibilizada.

Durante a análise da estrutura, foi observado que a coluna Código de autorização (na prática, tal coluna chama-se Dígito verificador), até o momento, não havia sido identificada. Na massa de dados, cada coluna apresenta valores numéricos; porém, utilizando-se a tabela de

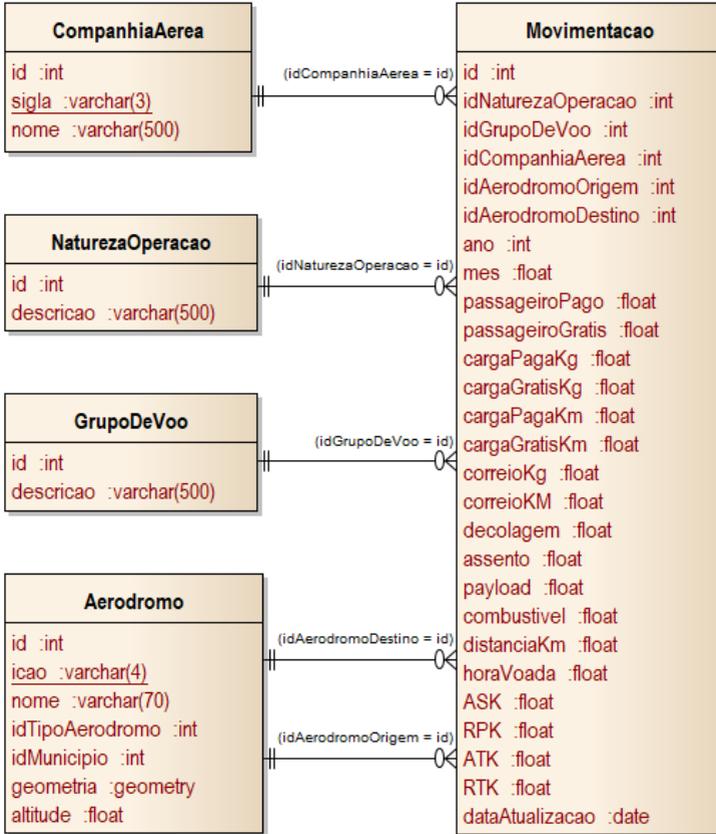


Figura 14 – Modelo estrutural do banco de dados para movimentação.  
 Fonte: O autor (2018).

descrição do código de autorização, disponibilizada juntamente com os dados, pôde-se identificar a descrição para cada valor nesse conjunto de dados. Assim, foi possível caracterizar o voo em 11 tipos diferentes.

Na análise de conteúdo da coluna Tipo de linha, identificou-se que esta estava diretamente ligada à estrutura de Natureza da operação (estrutura inicialmente estabelecida para dados estatísticos e HOTRAN). Para manter o alinhamento entre os dados das estruturas que possuem natureza da operação, foi realizada uma conversão nos possíveis valores dessa coluna. Os tipos de linha Especial, Sub-Regional, Nacional e Regional foram convertidos para Doméstico; os demais foram mantidos

no mesmo formato.

Os dados de partida e chegada previstas possuem informação de data e hora, enquanto que as colunas Partida e Chegada realizada possuem apenas data e hora se o voo estiver com valor na coluna Situação diferente de cancelado. Foi observado que, na grande maioria dos registros (90%), quando o voo está como cancelado, consta justificativa relacionada a "problemas técnicos" ou mesmo "autorizado a operar".

Para manter o conjunto de dados de VRA, a estrutura proposta é similar à disponibilizada pela ANAC, entretanto, os dados de companhia aérea, aeródromos (origem e destino), justificativa, código de autorização e tipo de linha (convertido para natureza de operação) são elementos que possuem correlação com outras estruturas; desse modo, tais dados farão apenas referência a tais tabelas. Na figura 15, é possível visualizar a relação entre as estruturas propostas para os dados de VRA.

### 4.3 CRIAÇÃO DA ESTRUTURA DE BANCO DE DADOS

Após coletar, estruturar e padronizar os dados, foi iniciado o procedimento de construção da base de dados propriamente dita. Para o armazenamento e a disponibilização dos dados coletados foi utilizado o Microsoft SQLServer 2016.

O conjunto de dados coletados correlaciona-se por meio de estruturas de tabelas auxiliares que permitem sua interligação entre os dados. Por isso, para a construção da base de dados, foi estabelecida uma estrutura que divide a base em *schemas*<sup>1</sup> por área de conhecimento e registro da estrutura, conforme segue:

- **dbo:** contempla uma estrutura de tabelas que dispõe de informações cadastrais e operações acerca do transporte aeroviário nacional. Deste modo, nesse *schema*, está disponível toda a estrutura de tabelas acerca do negócio de transporte aeroviário nacional; e
- **metadados:** contempla uma estrutura para o armazenamento dos metadados acerca das tabelas existentes na base de dados. Nesse conjunto de tabelas, são armazenados descritivos sobre a estrutura da base de dados, mantendo registrado, na base, um informativo sobre ela mesma.

---

<sup>1</sup>Coleções de objetos (tabelas) dentro de determinado banco de dados que organizam vários aspectos e são importantes para a segmentação da segurança, facilitando a administração dos objetos e dos dados.

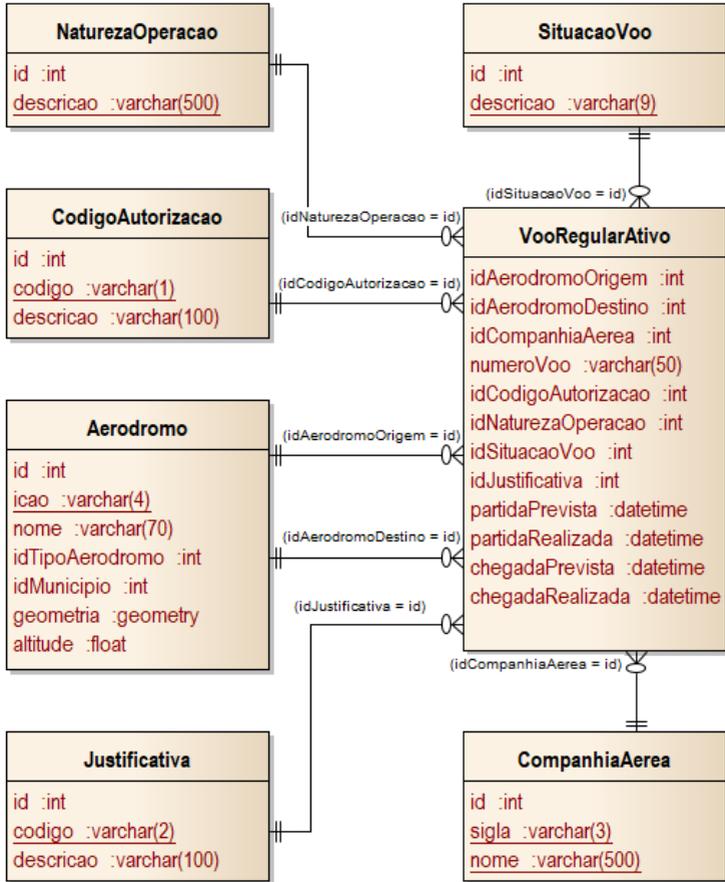


Figura 15 – Modelo estrutural do banco de dados para voo regular ativo.

Fonte: O autor (2018).

No *schema* dbo, estão centralizadas as informações acerca do transporte aéreo nacional propriamente ditas. Durante o processo de construção dessa base de dados, foram utilizadas as estruturas de dados estabelecidas nas atividades de análise da estrutura dos dados, bem como as regras estabelecidas na análise do conteúdo. Para cada tabela, foram estabelecidas chaves primárias (por meio de um identificador único na tabela) e, em alguns casos, chaves estrangeiras (coluna que

identifica o registro em outra tabela). Na figura 16, é apresentada, a estrutura de tabelas referente ao *schema* dbo, onde estão as tabelas das áreas de conhecimento. Nesse modelo, os elementos em verde representam, de forma direta, os dados coletados; já os elementos em amarelo representam os dados de apoio aos dados coletados.

Durante o processo de criação da base, foi utilizado um padrão de nomenclatura para tabelas e colunas. Para as tabelas, foi definido que seriam nomeadas com base na principal função dos dados armazenados. Além disso, os nomes das tabelas seriam escritos sempre com a primeira letra em maiúscula, seguida de caracteres minúsculos. Em caso de necessidade de utilização de nomes compostos, a separação foi realizada com a primeira letra no segundo nome em maiúscula e o restante das letras em minúsculas. Para as colunas, foi adotado um padrão em que todas as letras seriam minúsculas, com exceção de siglas, escritas com todas as letras em maiúsculas. Nos casos de colunas compostas por mais de uma palavra, foi adotado o mesmo padrão de separação das tabelas.

Para manter, na própria base de dados, as informações descritivas sobre ela, foi estabelecida uma estrutura de tabelas que permitisse o armazenamento dos registros dessas informações. Assim, toda tabela existente na base conta com um registro de sua descrição e fonte primária dos dados. Ademais, para o detalhamento das tabelas, existe uma estrutura para o registro de informações a respeito das colunas, tais como descrição, tipos de dados e tamanho das informações (número de caracteres). A figura 17 apresenta como foram estruturadas as tabelas do *schema* referente aos metadados.

Nessa estrutura, foi possível separar as tabelas de dados da estrutura dos dados sobre os dados (metadados), permitindo que os usuários da base tenham acesso a todo momento tanto aos dados quanto às informações sobre a base de dados.

#### 4.4 REGISTRO DE METADADOS

Com a estrutura criada, foram registrados, para cada tabela e coluna, os metadados na própria base de dados. No contexto deste trabalho, buscou-se estabelecer um padrão para as informações das estruturas da base de dados, no qual tais informações detalham o que é cada elemento. Para tabelas, são registrados nome, nome físico (utilizado na base de dados), descrição e fontes; para colunas, constam nome, nome físico, tipo de dado e tamanho.



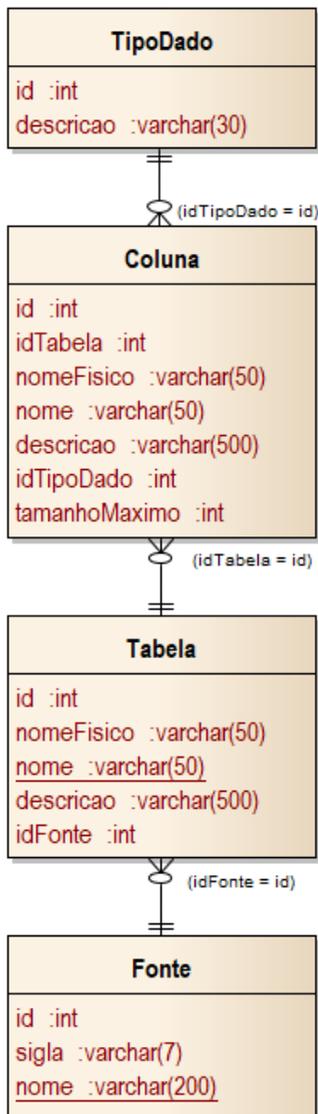


Figura 17 – Modelo para o *schema* metadado.  
Fonte: O autor (2018).

O processo de registro de metadados iniciou-se com informações de fontes de dados. A tabela de meta-dados requer, para cada fonte de dados, uma sigla e um nome. O conjunto de dados que a estrutura contempla possui registrados os órgãos responsáveis por cada informação. Assim, foram armazenados nessa estrutura os dados do IBGE e da ANAC.

A estrutura referente às tabelas registra o nome propriamente dito da tabela no banco de dados, o nome do conjunto de dados que a tabela armazena e um descritivo sobre a tabela. Na tabela 15, é possível visualizar um exemplo dos dados registrados na estrutura que documenta as tabelas.

Antes de iniciar o cadastro das informações de colunas, foi necessário cadastrar os tipos de dados que cada coluna poderia assumir. Em seguida, para os dados das colunas, o processo foi similar ao executado para as tabelas, no qual, para cada coluna, registrou-se o nome físico (nome usado na base de dados), o nome que a coluna representa, a descrição sobre o dado, o tipo de dado e o tamanho máximo para o dado.

#### 4.5 INSERÇÃO DOS REGISTROS

O procedimento de importação foi realizado carregando os dados dos arquivos coletados e posteriormente padronizados na base de dados criada. A atividade iniciou-se pelos dados de municípios, seguidos de dados de aeródromos, dados estatísticos (movimentação), horário de transporte e voo regular ativo. Para municípios, foi realizada importação dos dados sem reestruturação, isso porque a base criada tem o mesmo formato adotado pelo IBGE. A partir dessa atividade, as planilhas com informações de unidade federativa e municípios foram importadas diretamente para a base de dados. Os indicadores do IBGE para municípios foram inseridos manualmente na base – foram inseridos nove indicadores. Para os dados de indicadores dos municípios propriamente ditos, os dados foram importados em uma tabela temporária na base de dados. Após isso, os indicadores foram convertidos em identificadores da tabela Indicador, para que fosse possível transferir os dados para a estrutura final. Ao finalizar a transferência, a tabela temporária gerada foi removida da base de dados.

Para os dados de aeródromos, utilizou-se a planilha compilada durante a atividade de análise da estrutura dos dados, na qual estão contidos todos os aeródromos públicos e privados. O processo iniciou-

se com a importação dos dados para uma tabela temporária, com o mesmo formato disponibilizado pela ANAC. A partir dessa tabela, foram extraídos os dados de tipo de pavimentos por meio de uma consulta por valores únicos nas colunas de pavimentos da pista 1, da pista 2 e da pista 3. Ao finalizar os dados de pistas, foi realizado o cadastro dos aeródromos em si, atividade executada de forma direta, pois os municípios já haviam sido padronizados em conformidade com o IBGE. Com as colunas latitude e longitude convertidas para o formato geodésico decimal, utilizou-se a função do banco de dados que converte o dado no formato *geometry* (formato específico para elementos espaciais). Os dados de pistas foram importados fazendo-se uma relação do código gerado para aeródromo com a tabela temporária com dados de pistas. Após a importação dessas informações, a tabela temporária de aeródromos foi removida da base de dados.

Para os dados de estatísticos, horário de transporte e voo regular ativo, os dados das tabelas auxiliares – tais como companhia aérea, natureza da operação, grupo de voo, situação, código de autorização e justificativa – foram inseridos utilizando-se os dados tratados na atividade de análise do conteúdo. Os dados brutos (dados estatísticos, horário de transporte e voo regular ativo) foram importados por meio de um *script* (criado com a linguagem de programação Java) para uma estrutura temporária e, em seguida, transferidos (por meio do próprio *script*) para a estrutura de tabelas definidas.

#### 4.6 VALIDAÇÃO DOS DADOS

A validação da base de dados foi realizada em nível de tabela e seus registros, buscando aderência entre as fontes de dados e os registros importados para a base criada. Apesar de as informações terem sido reestruturadas e separadas em várias tabelas, cada uma dessas estruturas teve seus dados verificados, confrontando, dessa forma, todo o conteúdo importado para a base com as planilhas disponibilizadas pelas fontes de dados.

Para os dados de municípios, foi realizada a validação, em três níveis, das informações extraídas da planilha de indicadores disponibilizada pelo IBGE. O primeiro nível foi a conferência de quantitativos de registros importados para a base de dados em relação ao total de registros únicos existentes na planilha do IBGE. Para UF e municípios, a relação de comparação ocorreu de forma direta, pois como citado, foram removidos os dados duplicados e comparados com a base. Para os

indicadores, foi necessário contabilizar quantos registros existiam para cada ano e indicador, permitindo a comparação da base com a planilha. Outra validação realizada foi por meio da conferência de alguns registros em relação à base de dados. Para a verificação em nível de registro, determinou-se uma amostra de 35 registros por ano (total de 140 registros) para validação de forma visual. Para finalizar a verificação dos registros de municípios, foi realizada uma avaliação por meio da geração de gráficos anuais por indicador, confrontando a base de dados com a planilha. Na figura 18, consta um gráfico com o comparativo entre os totais anuais de todos os indicadores do IBGE.

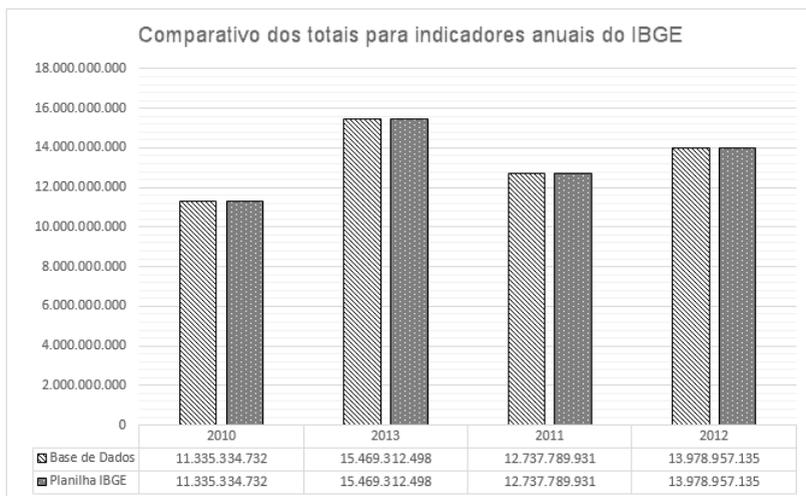


Figura 18 – Comparativo dos totais para indicadores anuais do IBGE.  
Fonte: O autor (2018).

Para a validação dos dados cadastrais de aeródromos, foram realizados dois tipos de procedimentos: verificação visual por amostra e verificação de totais de registros. Na verificação visual, foram selecionados aleatoriamente 70 aeródromos públicos e 70 aeródromos privados nas planilhas da ANAC. Após a seleção dos aeródromos a serem homologados na base de dados, comparou-se os dados da planilha com os dados importados na base. Como pode ser visualizado na figura 19, outra validação realizada foi por meio da verificação do total de aeródromos da planilha em relação ao total de aeródromos importados na base de dados.

No conjunto de dados que caracterizam um aeródromo, existem

The image shows a Microsoft Excel spreadsheet titled 'ANAC\_Aerodromos [Modo de Compatibilidade] - Excel'. The spreadsheet displays a list of airports with columns for ID, name, type, and location. A red box highlights the total number of records, '2.544 Registros'. Below the table, a green bar indicates '2544 row(s) returned (execu...' and 'CONTAGEM: 2544'. A SQL query window is overlaid on the spreadsheet, showing the following query:

```

SELECT
  id,
  icao,
  nome,
  idTipoAerodromo,
  idMunicipio,
  geometria,
  altitude
FROM dbo.Aerodromo
where idTipoAerodromo

```

ID	ICAO	NOME	Tipo	Município
2522	SIZB	Aeródromo Privado	Vale do Tucaná	Castanheira
2523	SVVL	Aeródromo Privado	Vale Europeu	Guaramirim
2524	SSVX	Aeródromo Privado	Valeu Boi Leilões	Rio Maria
2525	SIXS	Aeródromo Privado	Veracei	Belmonte
2526	SWAJ	Aeródromo Privado	Viatec Aviação Agrícola	Santa Vitória do Palm
2527	SWJP	Aeródromo Privado	Vila Bittencourt	Japurá
2528	SIBC	Aeródromo Privado	Vila de Tocós	Campos dos Goytacá
2529	SNNV	Aeródromo Privado	Vila Nova	Joinville
2530	SJVP	Aeródromo Privado	Vila Pitinga	Presidente Figueired
2531	SDFT	Aeródromo Privado	Virgolino de Oliveira	Itapira
2532	SJOT	Aeródromo Privado	Vista Alegre	Araguanã
2533	SDVY	Aeródromo Privado	Vitálii	Avanhandava
2534	SSTJ	Aeródromo Privado	Walter Ewald Siegel	Trombudo Central
2535	SDZK	Aeródromo Privado	Waphuta	Alto Alegre
2536	SJNE	Aeródromo Privado	Wapum	Caracará
2537	SDZN	Aeródromo Privado	Warogarem	Uiramutã
2538	SDZO	Aeródromo Privado	Waromada	Uiramutã
2539	SWJP	Aeródromo Privado	Warpol	São Desidério
2540	SJNF	Aeródromo Privado	Willimon	Uiramutã
2541	SNWR	Aeródromo Privado	Wilma Rebelo	Portel
2542	SJNZ	Aeródromo Privado	Xanadu	Fernandes Pinheiro
2543	SJNG	Aeródromo Privado	Xidea	Alto Alegre
2544	SJPO	Aeródromo Privado	Xingó	Piranhas
2545	SSZD	Aeródromo Privado	Xingó	Porto Velho

Figura 19 – Validação de dados cadastrais de aeródromos.

Fonte: O autor (2018).

os dados de regra de operação e pistas de pouso e decolagem. No processo citado para os dados de aeródromos, já considerou-se a validação dos dados de regras de operação. Para os dados de pista, além da verificação por meio da amostra, foi necessário realizar uma verificação dos totais de registros. Como pode ser visualizado na figura 20, para os dados de pistas, foi criado um conjunto de fórmulas na planilha de dados cadastrais de aeródromos, utilizando a ferramenta Microsoft Excel, o que permitiu realizar a conferência dos totais de registros para os dados de pistas.

Para os dados estatísticos (movimentação) e voo regular ativo, o processo de validação foi realizado por meio da conferência dos totais de registros importados na base de dados em relação aos dados disponibilizados pela fonte. Ainda foram gerados alguns gráficos com dados

	T	U	V	W	X	Y	Z	AA
1	PISTA3_DESIGNADOR	PISTA3_COMPRIMENTO	PISTA3_LARGURA	PISTA3_RESISTENCIA	PISTA3_SUPERFICIE			
2								
3								
4								
5								
6								
7						Total		2564
8						Pista 1		2544
9						Pista 2		19
10						Pista 3		1

Figura 20 – Validação dos dados de pistas.

Fonte: O autor (2018).

da base de dados e das planilhas disponibilizadas, os quais permitiram a verificação visual dos dados. Além desses elementos para validação, os dados foram conferidos de forma amostral: 100 registros para dados estatísticos e 120 registros para voo regular ativo.

O conjunto de dados de horário de transporte foi homologado utilizando-se as técnicas já citadas para os outros dados (conferência de registros, gráficos e conferência de totais), porém, foram incluídas algumas análises por meio de tabelas com diferenças entre totalizadores por coluna. Para tanto, foram desenvolvidas tabelas com dimensões por companhia aérea em relação ao total de registros, aeródromos em relação à quantidade de voos, equipamentos (aeronaves) em relação aos horários disponíveis e natureza da operação em relação à frequência e à quantidade de assentos ofertados. Na tabela 16, é possível visualizar um comparativo para cada natureza da operação e as variáveis de frequência semanal e quantidade de assentos.

Finalizada a validação, foi iniciado o processo de desenvolvimento de um protótipo de sistema para a disponibilização dos dados coletados e inseridos na base de dados. No próximo capítulo, tal sis-

tema será apresentado, juntamente com suas funcionalidades.

#### 4.7 DESENVOLVIMENTO DE SISTEMA PARA USO DOS DADOS

No presente trabalho, foi apresentado o método para construção de bases de dados, bem como sua aplicação para a construção de uma base de dados. O resultado dessa atividade foi a disponibilização de uma base de dados composta de informações públicas sobre o setor aeroportuário nacional.

A última atividade descrita no método proposto é denominada "Desenvolver sistema para uso dos dados"; portanto, os dados foram disponibilizados em um Sistema de Gerenciamento de Banco de Dados (SGBD), Microsoft SQLServer 2016. Assim, para acessar os dados da base disponibilizada, o usuário deve possuir conhecimentos de linguagem de programação SQL. Levando em conta tal contexto, confirmou-se a importância da existência de um sistema computacional que interprete e ofereça diferentes visões para as informações contidas na base, facilitando e apoiando o uso dos dados por parte dos especialistas da área de transportes. A partir das necessidades de uso do sistema, foram estabelecidos requisitos macro, sendo eles:

- apresentação de dados tabelados;
- apresentação de dados em gráficos;
- apresentação de dados georreferenciados;
- criação de mapas temáticos;
- visualização de origens e destinos de um aeroporto;
- exportação dos dados;
- atualização dos dados existentes; e
- inclusão de novos dados.

Após a identificação dos requisitos, estes passaram a ser implementados, sendo traduzidos para uma forma passível de execução por computadores. Com isso, foi desenvolvido um protótipo de sistema utilizando a linguagem de programação Java, bem como as bibliotecas GeoTools (para a exibição elementos geográficos) e JFreeChart (para a exibição de gráficos). Essas soluções tecnológicas foram

adotadas para a construção do protótipo pois são livres de licença de uso e oferecem extensa documentação. Para garantir que o protótipo de sistema tivesse uma apresentação adequada, foram utilizados, nas interfaces, ícones livres de licença de uso retirados do *site* Flaticon (<http://www.flaticon.com>). Como resultado tem-se um protótipo de sistema, denominado SIDDA (Sistema de Integração e Disponibilização de Dados Aeroportuários), que permite a manutenção e visualização dos dados contidos na base de dados.

#### **4.7.1 Sistema de Integração e Disponibilização de Dados Aeroportuários**

O SIDDA possui cinco módulos dos quais utilizam os dados disponíveis na base de dados e os apresentam por meio de mapas, indicadores e gráficos. Na figura 21, é possível visualizar os dados utilizados por cada um dos módulos do protótipo.

A interface principal tem a função de possibilitar acesso aos demais módulos do protótipo, bem como, é responsável pela apresentação dos dados em um sistema com características de SIG (Sistema de Informações Geográficas). No sistema de informações geográficas que integra o módulo, estão disponíveis as funcionalidades básicas e de características de um sistema SIG, sendo elas: *zoom in*, *zoom out*, *zoom* inicial, arrastar, informações sobre os elementos geográficos, rosa dos ventos e posição do cursor no mapa. Existe ainda uma opção que permite gerenciar as camadas geográficas disponíveis, sendo possível ordenar a sobreposição das camadas, ativar ou desativar uma camada e visualizar dados da camada geográfica.

Para a disponibilização dos dados espaciais, foram utilizados dados públicos advindos do Portal de Dados Abertos que apresentam a forma geoespacial de países, estados e municípios. As camadas geográficas que representam aeródromos públicos e privados foram geradas utilizando-se os dados de coordenadas disponíveis na própria base de dados criada. Na figura 22, é possível visualizar a interface principal do protótipo.

Como a interface principal de acesso aos demais módulos do protótipo, nele, foram disponibilizados tanto ícones (principais rotas entre aeródromos, mapas e indicadores temáticos) quanto uma barra de menu para acesso a tais módulos.

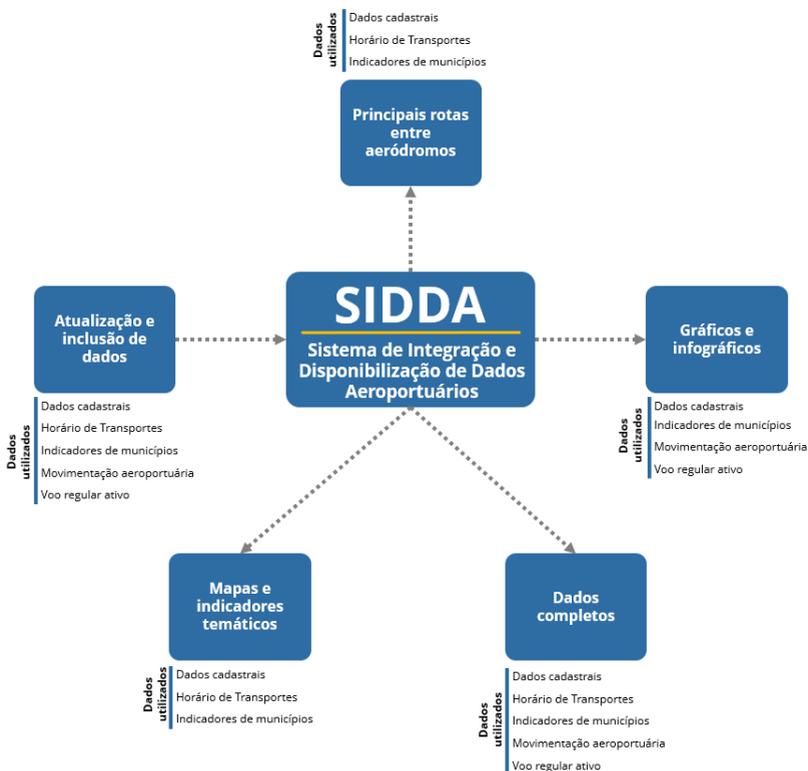


Figura 21 – Módulos do protótipo e dados utilizados.  
Fonte: O autor (2018).

#### 4.7.1.1 Principais rotas entre aeródromos

O módulo de principais rotas entre aeródromos permite a visualização das rotas aéreas autorizadas de determinado aeródromo e de informações detalhadas acerca dos voos de origem e destino do aeródromo selecionado. O resultado dessa funcionalidade é a representação de origens e destinos planejados por meio de linhas de desejo para o modal aéreo. Para visualizar as rotas de um aeródromo, é necessário realizar a seleção de alguns parâmetros (Figura 23), sendo eles:

- **Aeródromo:** aeródromo em que se deseja visualizar as rotas.
- **Sentido:** sentido dos voos em relação ao aeródromo selecionado:

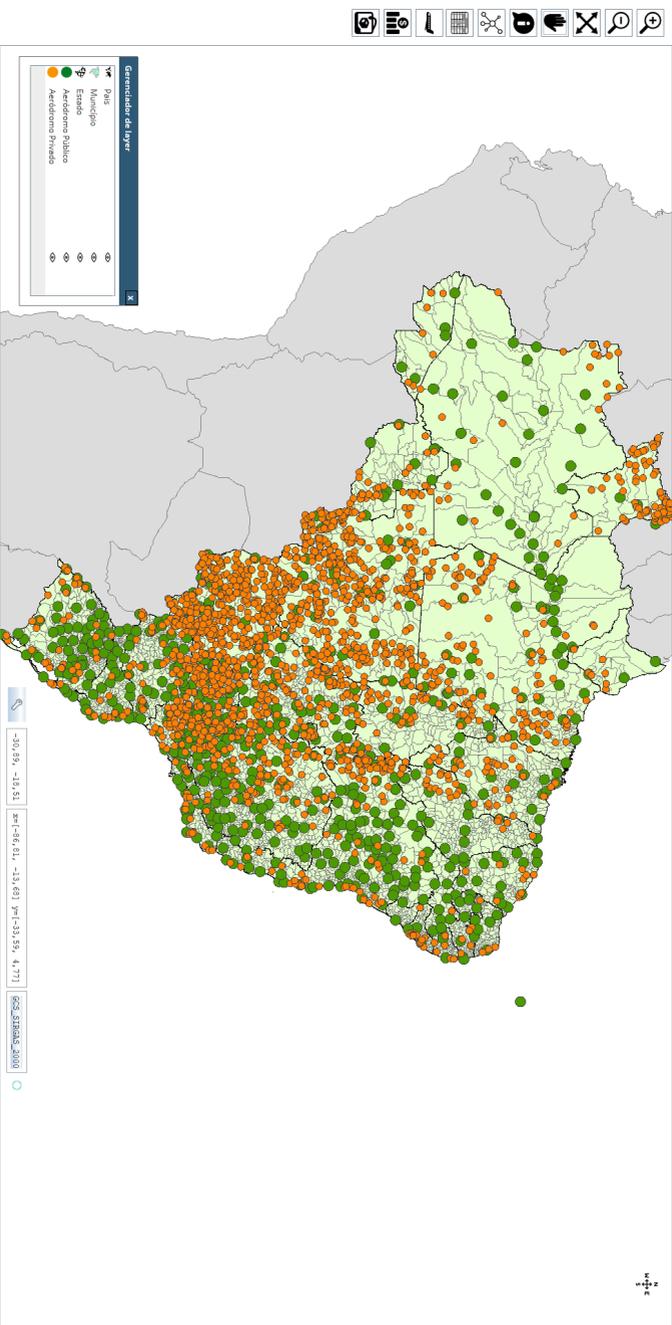


Figura 22 – Interface principal do protótipo.  
Fonte: O autor (2018).

SIDDA - Sistema de Integração e Disponibilização de Dados Aeroportuários

### Principais rotas entre aeródromos

Aeródromo: SBCT

Sentido:  origem  destino  ambos

Quantidade de pares: 10

Quantidade de faixas: 5

Cor para aeródromos:

Cor para rotas:

Executar

Figura 23 – Seleção de rotas por aeródromo.

Fonte: O autor (2018).

- **Origem:** aeródromos que possuem voos cujo aeródromo de origem é o selecionado;
  - **Destino:** aeródromos que possuem voos cujo aeródromo de destino é o selecionado; e
  - **Ambos:** aeródromos que possuem voos cujo aeródromo de origem ou destino é o selecionado.
- **Quantidade de pares:** número máximo de aeródromos com voos regulares para o aeródromo selecionado.
  - **Quantidade de faixas:** número máximo de classes para configuração das linhas (ligação entre aeródromos).
  - **Cor para aeródromos:** cor para representação dos aeródromos no mapa.
  - **Cor para rotas:** cor para representação das linhas no mapa.

Na configuração das principais rotas, a quantidade máxima (apresentada por *default*) de aeródromos com rotas para o aeródromo selecionado é de 10 rotas; já a quantidade de faixas apresentadas por *default* é de sete faixas. A cor *default* para aeródromos é vermelha e, para rotas, azul.

Para a apresentação do módulo, neste trabalho, foi selecionado o aeródromo de Curitiba (SBCT); em relação aos demais parâmetros, foram utilizados aqueles disponibilizados por *default* no protótipo. Após configurar dos parâmetros e clicar em executar, são utilizados dados de HOTRAN, dados socioeconômicos do IBGE e dados cadastrais de aeródromos para a criação do mapa proposto. Os dados de HOTRAN são utilizados para a definição das rotas, bem como para o cálculo da frequência com que os voos ocorrem. Do conjunto de dados socioeconômicos, são utilizados os PIBs dos municípios que estão na mesma microrregião do município que sedia o aeródromo e que possuem voos regulares para o aeródromo objeto de estudo. Na figura 24, é possível visualizar um mapa resultante dessa funcionalidade, cujo aeródromo selecionado foi o de Curitiba, conforme mencionado.

Para complementar o mapa, é apresentada uma interface com um detalhamento dos resultados nele disponíveis, a saber, indicadores sobre o conjunto de rotas (total de voos, menor frequência e maior frequência) e gráfico com as frequências de voos entre os aeródromos de origem ou destino em relação ao aeródromo selecionado (no caso, Curitiba). Tais dados são representados em um *grid* abaixo do gráfico. Por fim, é disponibilizado ainda um *grid* de dados com o PIB (em ordem decrescente) dos municípios que estão na microrregião do município sede dos aeródromos e as faixas nas eles se enquadram (caracterizadas por cor).

#### 4.7.1.2 Mapas e indicadores temáticos

O módulo de mapas temáticos dispõe de funcionalidades que permitem a construção de mapas por meio de dados oriundos da ANAC e do IBGE. A elaboração do mapa é guiada pelo protótipo para que a sua execução se dê de forma simples. Nesse contexto, tal módulo permite a representação de um dado alfanumérico (referente a um registro no tempo) de forma geográfica. Para isso, os dados alfanuméricos são relacionados com os dados espaciais de aeródromos, municípios e estados.

No contexto do protótipo de sistema, a funcionalidade de mapas temáticos de aeródromos é o meio para a criação de mapas e indicadores temáticos por meio da classificação dos aeródromos por seu volume de movimentação. Para construir um tema sob a camada de aeródromos, é necessário realizar a seleção de alguns parâmetros (figura 25), sendo eles:

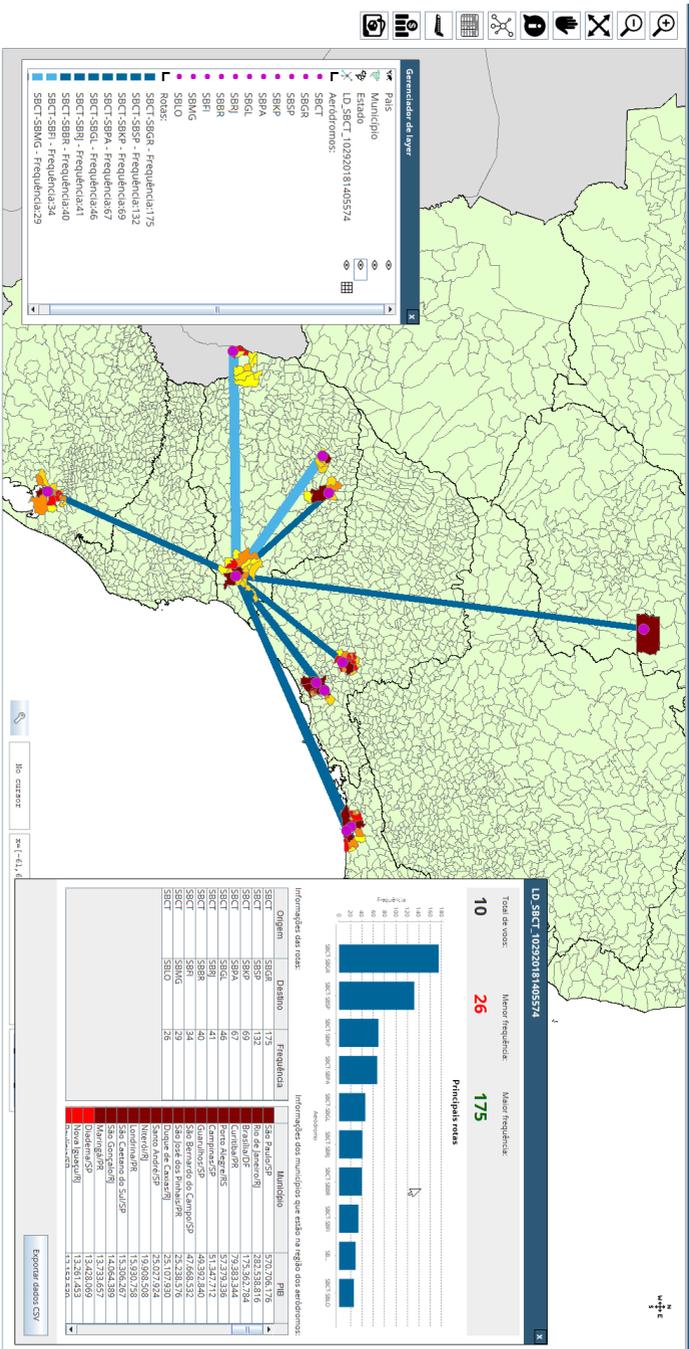


Figura 24 – Visualização das principais rotas do aeródromo (linhas de desejo).  
 Fonte: O autor (2018).

- **Operação:** permite definir o tipo operacional, sendo eles:
  - **Passageiro:** número de passageiros movimentados no aeródromo;
  - **Aeronave:** número de aeronaves movimentadas no aeródromo;
  - **Mala postal:** volume de malas postais movimentadas no aeródromo;
  - **Carga aérea:** volume de cargas aéreas movimentadas no aeródromo.
- **Quantidade de aeródromos:** número máximo de aeródromos com as maiores movimentações registradas.
- **Ano:** ano de referência para o dado de movimentação.
- **Cor para aeródromos:** cor para representação dos aeródromos no mapa.

SIDDA - Sistema de Integração e Disponibilização de Dados Aeroportuários

### Mapa e indicadores temáticos

aeródromos

Operação

Passageiro
  Aeronave
  Mala postal
  Carga Aérea

Quantidade de aeródromos:

Cor para aeródromos:

Ano:

Executar

Figura 25 – Criação de mapa e indicadores temáticos de aeródromos por movimentação.

Fonte: O autor (2018).

Para facilitar a interação entre usuário e o protótipo de sistema, a funcionalidade de temática de aeródromos apresenta os parâmetros previamente definidos. Após configurar os parâmetros, basta clicar no botão Executar para que o protótipo apresente um mapa temático e informações detalhadas sobre ele. A figura 26 apresenta o resultado desse módulo utilizando-se como parâmetros de movimentação de passageiros: quantidade de aeródromos 15, ano 2015 e cor vermelha para aeródromos.

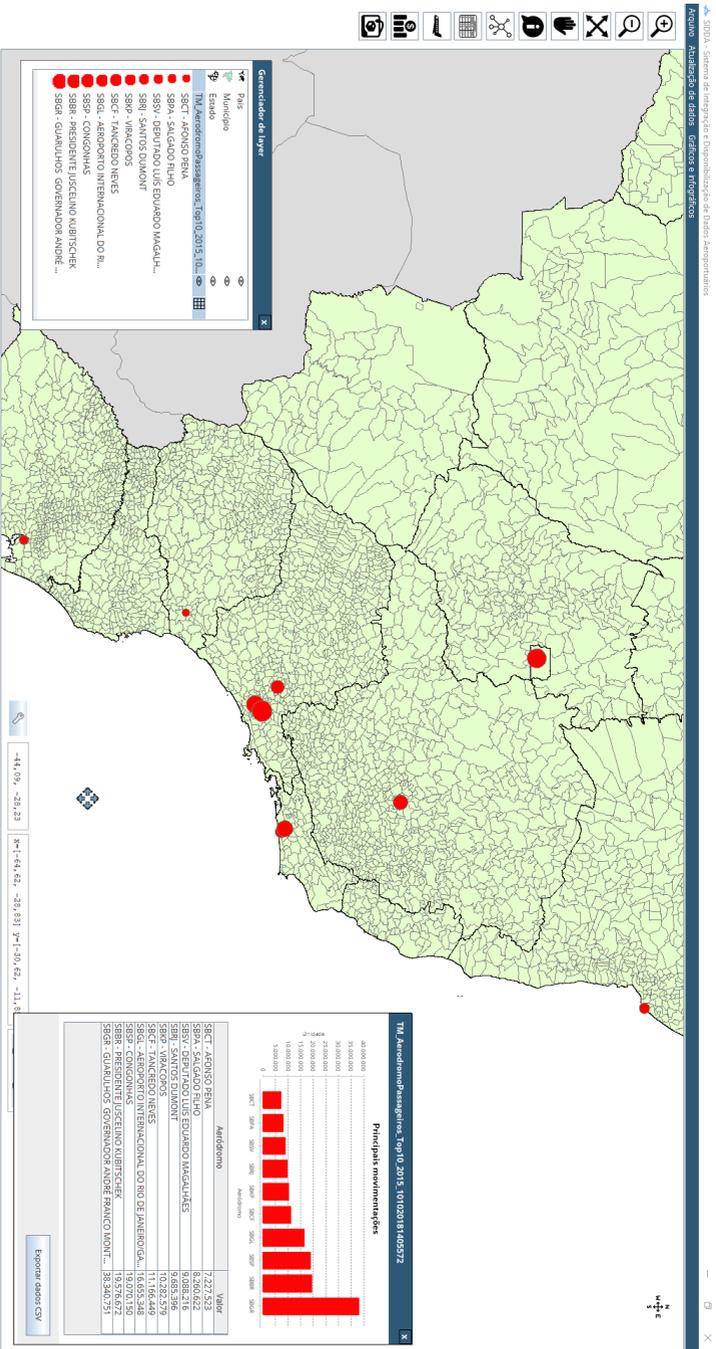


Figura 26 – Mapa e indicadores temáticos de aeródromos por movimentação.  
 Fonte: O autor (2018).

Os mapas e indicadores temáticos de aeródromos utilizam dados de movimentação de passageiros, aeronaves, carga aérea e mala postal; para isso, são consultados os dados da ANAC. Como complemento ao mapa, os registros de movimentação de cada um dos aeródromos que fazem parte do tema criado são disponibilizados em gráfico e *grid* de dados.

Para o desenvolvimento da funcionalidade, foi necessário agrupar todos os registros de movimentação disponíveis para um aeródromo por meio de consultas à base, isso porque tais dados disponibilizam registros par a par de aeródromo de origem e destino, pois a representação do dado no mapa é realizada por meio de um valor único para aeródromo. Portanto, o agrupamento realizado soma toda a movimentação de um aeródromo quando ele é origem com toda a movimentação quando ele é destino. O valor total obtido nessa soma para o aeródromo é disponibilizado em mapa, gráfico e *grid* de dados.

No contexto do protótipo, a funcionalidade de mapas e indicadores temáticos de indicadores socioeconômicos é a forma de elaborar mapas por meio da classificação de estados ou municípios. Tais mapas e indicadores podem ser criados para um ano, com dados de PIB, população, PIB *per capita*, impostos, serviços, administração, agropecuária e total de impostos (bruto total).

Para a construção de um tema na funcionalidade, é necessário selecionar agrupamento (estado ou município), ano e um indicador socioeconômico; em seguida, basta clicar no botão Executar. Como resultado, o protótipo apresenta o mapa e indicadores temáticos com os dados do IBGE. Na figura 27, é possível visualizar a interface de configuração de mapa e indicadores temáticos de indicadores socioeconômicos.

Para a criação do mapa temático de indicadores socioeconômicos para municípios, utiliza-se os dados do IBGE do modo que estão dispostos, sem qualquer manipulação, apenas aplicando o filtro ano. Para demonstrar o resultado da funcionalidade, é apresentado um mapa para o agrupamento municípios, ano 2013 e indicador PIB, na figura 28. Além do mapa propriamente dito, é apresentado um *grid* com informações detalhadas sobre os indicadores selecionados para municípios.

Os mapas e indicadores temáticos utiliza os dados do IBGE de forma agrupada; como o tema é por estados e o IBGE faz disponibilização por municípios, o sistema agrupa os dados de municípios para seus respectivos estados. Para demonstrar o resultado da funcionalidade, é apresentado um mapa para o agrupamento estados, ano 2013 e indicador PIB, na figura 29. Além do mapa propriamente dito, é apresentado um *grid* com informações detalhadas sobre os indicadores



Figura 27 – Criação de mapa e indicadores temáticos para dados do IBGE.

Fonte: O autor (2018).

selecionados para estados.

#### 4.7.1.3 Gráficos e infográficos

O módulo de gráficos infográficos é responsável por apresentar dados movimentação, indicadores socioeconômicos e voo regular ativo

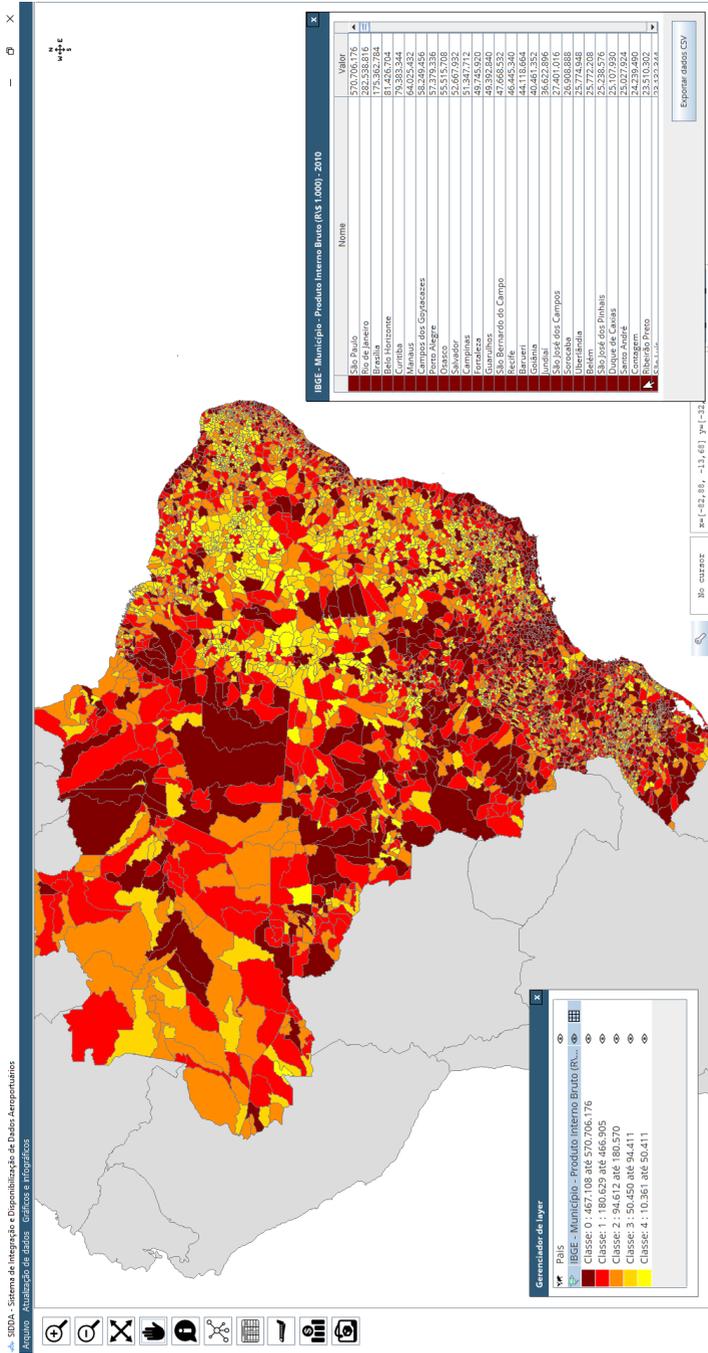


Figura 28 – Mapa e indicadores temáticos por municípios para dados do IBGE. Fonte: O autor (2018).



de forma gráfica. Esse formato de disponibilização dos dados oferece ao usuário uma visão gerencial acerca do cenário aeroportuário brasileiro, auxiliando em uma rápida interpretação das informações.

O conjunto de dados propostos para a visualização em gráficos é bastante extenso. Assim, para a construção deste módulo, foi identificada a necessidade de consolidar tais dados de forma otimizada a fim de obter-se melhor desempenho durante as consultas e a exibição dos gráficos.

Para os dados de movimentação, oriundos da ANAC, são apresentados: gráfico de evolução anual (apresenta o quantitativo total da movimentação para os aeródromos ao longo do ano), gráfico de participação por companhia aérea (conforme a movimentação total, apresenta o percentual movimentado para cada companhia aérea em determinado ano) e gráfico de *ranking* de movimentação (apresenta a movimentação dos 20 principais aeródromos em determinado ano).

Para a parte de movimentação, são apresentados gráficos para cada um dos tipos operacionais, sendo eles passageiro, aeronave, mala postal e carga aérea. Cada tipo operacional tem seus gráficos dispostos em uma aba. Na figura 30, é possível visualizar os dados de movimentação para os dados de passageiros.

Para os indicadores socioeconômicos, oriundos do IBGE, são apresentados: gráfico de evolução (apresenta os dados de PIB e PB *per capita* ao longo dos anos), gráfico de participação de indicadores (apresenta a participação dos valores de impostos por aérea em relação ao total de impostos para determinado ano) e gráfico de evolução da população (apresenta a quantidade de pessoas registradas ao longo dos anos). Na figura 31, é possível visualizar os gráficos propostos para os indicadores socioeconômicos.

Para os dados de voo regular ativo, oriundos da ANAC, são apresentados gráficos de hora pico (dispondo a quantidade de registros de voos organizados por cada faixa de horário em uma visão para todos os aeródromos cadastrados), situação de voo (apresenta os percentuais de voos cancelados e atrasados) e *ranking* de aeródromos (considerando os dados de partida, apresenta os 10 maiores aeródromos em número de partidas).

Em relação aos dados de voo regular ativo, pelo fato de apresentarem data e hora de cada voo, foi possível criar um gráfico com o volume de voos por horário. Esse tipo de dado permite ao planejador aeroportuário identificar gargalos na rede e, por conseguinte, pontos que merecem atenção em relação ao volume de aeronaves. Além disso, tal dado indica o nível de serviço dos aeródromos em relação ao cumpri-

Gráficos e Infográficos  
Movimentação

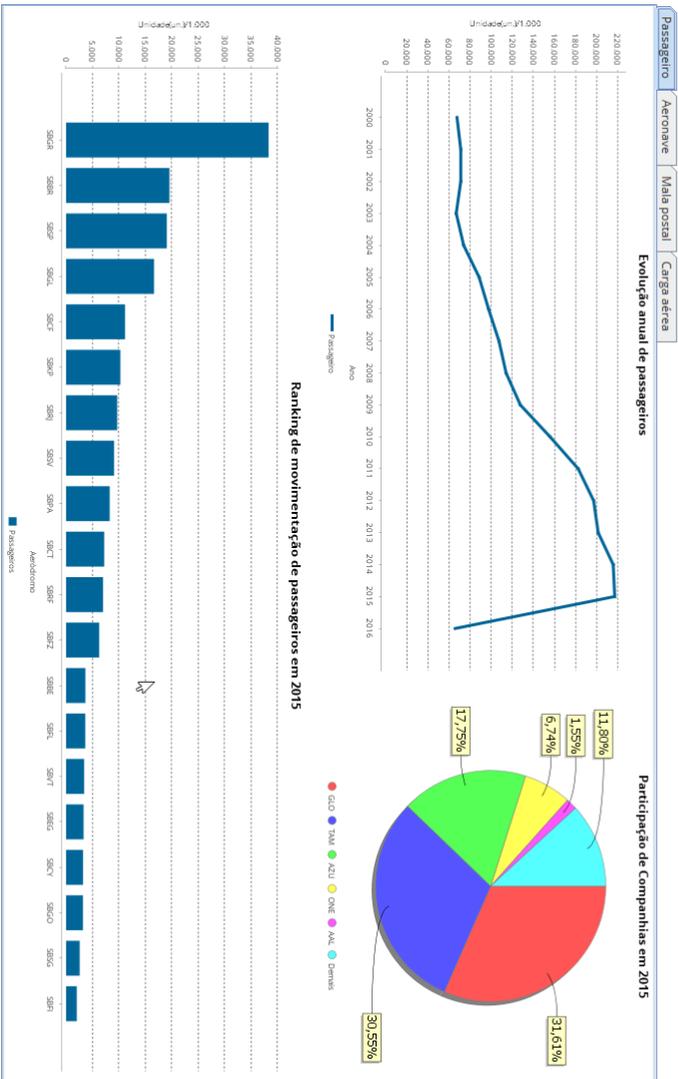
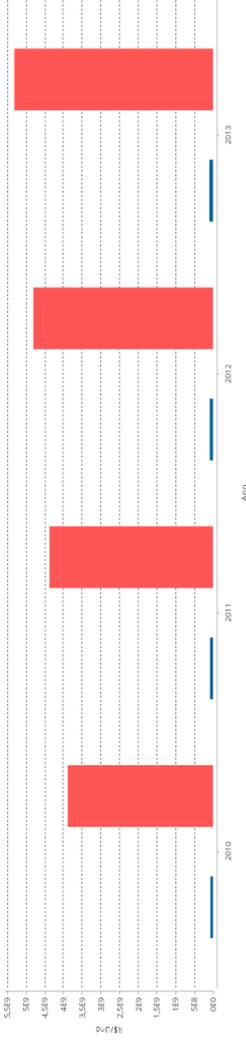


Figura 30 – Gráficos para dados de movimentação aeroportuária - passageiros.  
Fonte: O autor (2018).

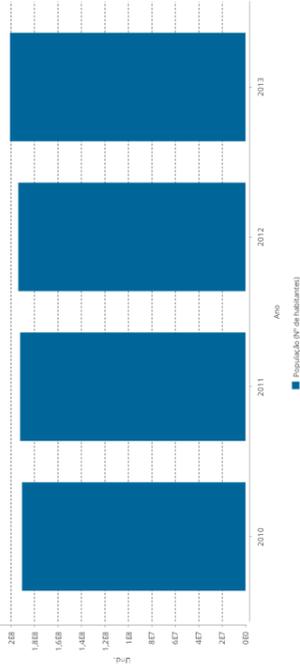
### Gráficos e infográficos

indicadores socioeconômicos

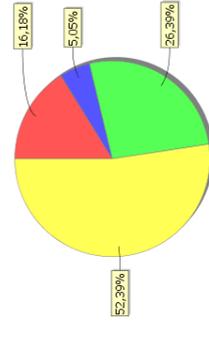
#### Evolução dos indicadores



#### População



#### Participação Indicadores



- Valor adicionado bruto da Administração, saúde e educação pública e atividades sociais, a preços correntes (R\$ 1.000)
- Valor adicionado bruto da Agricultura, a preços correntes (R\$ 1.000)
- Valor adicionado bruto dos Serviços, a preços correntes - excluindo Administração, saúde e educação I
- Valor adicionado bruto da indústria, a preços correntes (R\$ 1.000)

Fonte: IBGE

Figura 31 – Gráficos para dados do IBGE. Fonte: O autor (2018).

mento de horários (atrasos) e cumprimento de voos (cancelamentos).

Para oferecer ao usuário maior flexibilidade quanto ao uso dos dados de voo regular ativo, foi disponibilizado um filtro na interface, o qual permite estabelecer qual período dos dados se deseja visualizar. Na figura 32, é possível verificar a interface de gráficos para voo regular ativo.

#### 4.7.1.4 Dados completos

O módulo de dados completos tem como objetivo disponibilizar todos os dados existentes na base de dados por meio de uma grade de dados. Desse modo, o usuário não precisa ter conhecimentos específicos de linguagens de programação ou banco de dados para a visualização e o uso das informações contidas no protótipo de sistema.

Além de apresentar os dados da forma como estão armazenados no protótipo sistema, o módulo traz visões específicas deles para potencializar o modo de visualização dos dados. Assim, estão disponíveis as seguintes visões dos dados:

- **Listagem de aeródromos:** dados cadastrais de aeródromos existentes na base de dados, contendo: código identificador, código ICAO, nome e município em que está situado.
- **Listagem de aeródromos (completo):** dados cadastrais de aeródromos existentes na base de dados, contendo: código identificador, código ICAO, nome, município em que está situado, coordenadas (latitude e longitude) e informações de pista.
- **Listagem de estados:** dados cadastrais de estados existentes na base de dados, contendo: código identificador, nome e sigla.
- **Listagem de municípios:** dados cadastrais de municípios existentes na base de dados, contendo: código identificador, código IBGE, nome e unidade da federação.
- **Indicadores socioeconômicos (municípios):** indicadores socioeconômicos dos municípios existentes na base de dados, contendo: código identificador, código IBGE, nome, unidade da federação, ano, PIB, população, PIB *per capita*, impostos, serviços, administração, agropecuária e total de impostos (bruto total).
- **Indicadores socioeconômicos (estados):** indicadores socioeconômicos dos estados existentes na base de dados, contendo: uni-

X

SIDA - Sistema de Integração e Disponibilização de Dados Aeroportuários

**Gráficos e infográficos**  
Voo Regular Ativo

Filtro  
3/2016

Filtrar

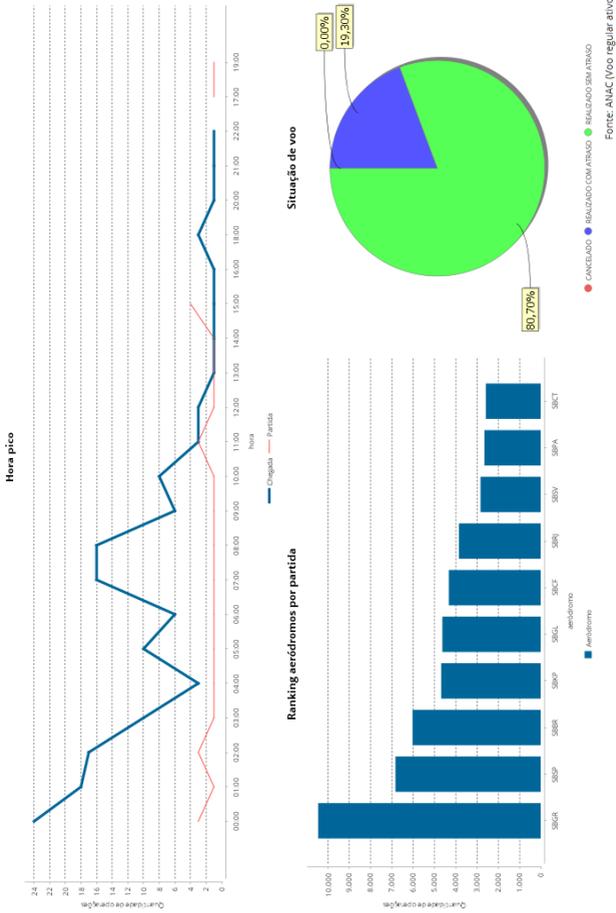


Figura 32 – Gráficos para dados VRA.  
Fonte: O autor (2018).

dade da federação, sigla, ano, PIB, população, impostos, serviços, administração, agropecuária e total de impostos (bruto total).

- **Movimentação (completo):** dados de movimentação aeroportuária completos existentes na base de dados, contendo: código ICAO do aeródromo de origem, nome do aeródromo de origem, código ICAO do aeródromo de destino, nome do aeródromo de destino, ano, natureza da operação, grupo de voo, passageiros, aeronaves, carga aérea, mala postal, combustível, distância voada, assentos, ASK, RPK, ATK, RTK, *payload* e horas voadas.
- **Movimentação por aeródromo:** dados de movimentação agregada por aeródromo existente na base de dados, contendo: código ICAO, nome do aeródromo, município, ano, natureza da operação, grupo de voo, passageiros, aeronaves, carga aérea, mala postal, combustível, distância voada, assentos, *payload* e horas voadas.
- **Movimentação comparativo por aeródromo:** dados de movimentação aeroportuária existentes na base de dados, comparando o total movimentado no ano mais recente em relação ao anterior, contendo: código ICAO, nome do aeródromo, município, ano, total de passageiros (ano A<sup>2</sup>), total de passageiros (ano B<sup>3</sup>), percentual de variação de passageiros, aeronaves (ano A), aeronaves (ano B), percentual de variação de aeronaves, carga aérea (ano A), carga aérea (ano B), percentual de variação para carga aérea, mala postal (ano A), mala postal (ano B) e percentual de variação para mala postal.
- **Horário de transporte (completo):** dados de Horário de transporte existentes na base de dados, contendo: data de coleta, código da empresa, nome da empresa, número do voo, equipamento, dias da semana, quantidade de assentos, número HOTRAN, data de solicitação, data de aprovação, data de vigência, natureza da operação, número da etapa, código ICAO da origem, código ICAO do destino, horário de partida e horário de chegada.
- **Horário de transporte (vigente):** dados de Horário de transporte recente, contendo: código da empresa, nome da empresa, número do voo, equipamento, dias da semana, quantidade de assentos, número HOTRAN, data de solicitação, data de aprovação,

---

<sup>2</sup>Representa o total movimentado para o tipo operação para o ano mais recente.

<sup>3</sup>Representa o total movimentado para o tipo operação para o ano anterior ao mais recente.

data de vigência, natureza da operação, número da etapa, código ICAO da origem, código ICAO do destino, horário de partida e horário de chegada.

- **Horário de transporte (frequência semanal):** dados de Horário de transporte agrupados por aeródromos, contendo: código ICAO do aeródromo, dias da semana, frequência semanal e quantidade de assentos.
- **Voo regular ativo:** dados de voo regular ativo existentes na base de dados, contendo: código da empresa, número do voo, código da autorização, tipo de linha, código ICAO do aeródromo de origem, código ICAO do aeródromo de destino, partida prevista, partida real, chegada prevista, chegada real, situação do voo e justificativa.
- **Voo regular ativo (atraso e cancelamento):** dados de voo regular ativo agrupados por situação dos voos, contendo: código ICAO do aeródromo, ano, mês, quantidade de voos cancelados, quantidade de voos realizados sem atrasos e quantidade de voos realizados com atrasos.
- **Voo regular ativo (hora pico):** dados de voo regular ativo agrupados por situação dos voos hora a hora, contendo: código ICAO do aeródromo, ano, mês, hora, quantidade de voos cancelados, quantidade de voos realizados sem atrasos e quantidade de voos realizados com atrasos.

Para alterar os dados a serem visualizados, bastar realizar a seleção da visão desejada na parte superior da interface e clicar no botão Carregar. Para todos os dados, é possível realizar ordenação (qualquer coluna) e aplicar filtros. Ainda, há a opção de exportar os dados em formato CSV, como pode ser verificado na figura 33.

#### 4.7.1.5 Atualização e inclusão de dados

O módulo de atualização é responsável por garantir que os dados existentes na base e no protótipo de sistema estejam atualizados conforme a disponibilidade em sua fonte. Na prática, esse módulo automatiza a etapa de atualização de dados do método proposto neste trabalho, isso para dados conhecidos e já estruturados.

## Dados completos ANAC - Lista de aeródromos

Carregar

ID	ICAO	Aeródromo	Município
406	SOSW	ILHA DAS FLORES	Ara Floresta DOeste
1162	SJTF	FAZENDA MEQUENS	Ara Floresta DOeste
1094	SJOG	ARIQUEMES	Ariqueemes
2428	SWTB	TABAJARA	Ariqueemes
1312	SJNC	FAZENDA CENTRINO	Cabixi
1890	SSRW	CACICAL	Caracai
2143	SWBO	FAZENDA PEROBAL	Corumbiara
874	SJZN	FAZENDA SANTA ANA	Corumbiara
878	SJZR	FAZENDA CAROLINA	Corumbiara
2400	SWRL	FAZENDA POUSO REDONDO	Corumbiara
2497	SWXC	FAZENDA TROPICAL	Corumbiara
2162	SWCQ	COSTA MARQUES	Costa Marques
48	SBGM	GUAJARÁMIRIM	Guajará-Mirim
2397	SWRG	RANCHO MARIA E TEREZA	Guajará-Mirim
537	SIBN	SAVANA	Jaru
547	SI0G	IRIAMAOS GONCALVES	Jaru

Total de registros: 2544

Exportar CSV

Figura 33 – Visualização de dados completos.  
Fonte: O autor (2018).

No processo de atualização, o protótipo utiliza os dados originais (disponibilizados pela fonte de dados) convertendo-os e reagrupando-os para todas as visões que os demandam. Cada conjunto de dados possui um método diferente para a sua obtenção devido ao modo como são disponibilizados. Sendo assim, para o processo de atualização, os seguintes dados são carregados automaticamente:

- **Dados cadastrais de aeródromos:** para a obtenção do dado, é acessado o *site* da ANAC e, na área de aeródromos, são coletados os arquivos que dispõem de informações de aeródromos públicos<sup>4</sup> e aeródromos privados<sup>5</sup>, que estão disponíveis em formato CSV. De posse desse material, é realizada uma varredura, sendo os dados carregados linha a linha para a base do protótipo de sistema. Toda vez que esse processo é executado, os dados existentes na base são substituídos pelos novos.
- **Horário de transporte:** para a obtenção do dado, é realizado um acesso direto ao serviço *web*<sup>6</sup>, sendo necessário informar como parâmetro a data desejada. Como retorno desse serviço *web*, são disponibilizados os dados tabulados; após a varredura, eles são inseridos na base de dados do protótipo de sistema. Os dados são incluídos na base de forma incremental, mantendo os registros anteriormente importados.
- **Movimentação:** para a obtenção do dado, é acessado o *site* da ANAC e, na área de estatísticas do setor aeroportuário, é coletado o arquivo que dispõe de toda a movimentação, o qual está disponível em Dados Estatísticos<sup>7</sup>. O arquivo é disponibilizado de forma compactada, portanto, é necessário descompactá-lo para ter acesso ao conteúdo, que, por sua vez, está em formato CSV. De posse do arquivo, é realizada uma varredura e carregada linha a linha para a base de dados do protótipo de sistema. Toda vez que o processo é executado, os dados existentes na base são substituídos pelos novos.

---

<sup>4</sup>Disponível em: <http://www.anac.gov.br/assuntos/setor-regulado/aerodromos/acesso-rapido/lista-de-aerodromos-publicos>. Acesso em: 7 mar. 2018.

<sup>5</sup>Disponível em: <http://www.anac.gov.br/assuntos/setor-regulado/aerodromos/acesso-rapido/lista-de-aerodromos-publicos>. Acesso em: 7 mar. 2018.

<sup>6</sup>Disponível em: [http://www2.anac.gov.br/hotran/hotran\\\_data.asp](http://www2.anac.gov.br/hotran/hotran\_data.asp). Acesso em: 7 mar. 2018.

<sup>7</sup>Disponível em: [http://www.anac.gov.br/assuntos/dados-e-estatisticas/dados-estatisticos/arquivos/Dados\\\_Estatisticos.zip](http://www.anac.gov.br/assuntos/dados-e-estatisticas/dados-estatisticos/arquivos/Dados\_Estatisticos.zip). Acesso em: 7 mar. 2018.

- **Municípios:** para a obtenção do dado, é acessado o FTP<sup>8</sup> do IBGE. É feita uma busca nas pastas à procura da mais atual (as pastas são identificadas por ano) e, após encontrar a mais recente, é coletado o arquivo cujo nome inicia-se com "PIB\_municipios". De posse do arquivo, é realizada uma varredura e carregada linha a linha para a base de dados do protótipo de sistema. Toda vez que o processo é executado, os dados existentes na base são substituídos pelos novos.
- **Voo regular ativo:** para a obtenção do dado, é feito um acesso direto ao serviço *web*<sup>9</sup>, sendo necessário informar como parâmetro a data desejada. Como retorno, é disponibilizado o dado tabulado. Após uma varredura, os dados são inseridos na base do protótipo de sistema. Tais dados são incluídos na base de forma incremental, mantendo os registros anteriormente importados.

Para cada dado a ser atualizado, é requerido um conjunto de informações inicial, indicando seu local de disponibilização. Após o acionamento do usuário (clique no botão Atualizar dados), o protótipo de sistema segue as etapas do método proposto (indicando em tempo de execução quais etapas foram executadas) para a atualização dos dados. Na figura 34, é possível visualizar a interface de atualização dos dados de HOTRAN.

Durante a atualização, o protótipo de sistema ainda apresenta um detalhamento sobre o processo em formato textual a fim de proporcionar ao usuário melhor entendimento sobre o que está sendo executado. Tal módulo é de suma importância, pois permite que usuários leigos mantenham o protótipo e a base de dados sempre atualizados.

Com a finalidade de permitir que os usuários incluam novos dados na base, é disponibilizado um módulo que permite a inserção de dados de forma orientada, seguindo o método proposto. Assim, com a simples seleção de um arquivo de dados (em formato CSV) e a configuração de informações básicas (como nome, fonte, descrição e características das colunas), é possível incluir um novo dado no protótipo de sistema. Na figura 35, é possível visualizar a interface elaborada para esse fim.

Quando da seleção de um arquivo no processo de inclusão de dados, é possível criar um relacionamento entre os dados de determinada

---

<sup>8</sup>Disponível em: [ftp://ftp.ibge.gov.br/Pib/\\_Municipios/](ftp://ftp.ibge.gov.br/Pib/_Municipios/). Acesso em: 7 mar. 2018.

<sup>9</sup>Disponível em: <https://sistemas.anac.gov.br/bav/FormConsultarVraInternet.do>. Acesso em: 7 mar. 2018.

## Atualização de dados de HOTRAN

Informações de HOTRAN

Endereço (URL):

Data inicial (mm/dd/aaaa):

\* Será incluído automaticamente o HOTRAN de 09/05/2018

100%

Ações executadas:

- Preparando estrutura de dados
- Preparando dados
- Coletando dados
- Verificando Companhia Aérea
- Verificando aeronave
- Verificando aeródromos
- Transformando dados
- Carregando dados
- Carregando frequência de rotas
- Carregando hora-pico
- Finalizando importação
- Importação concluída

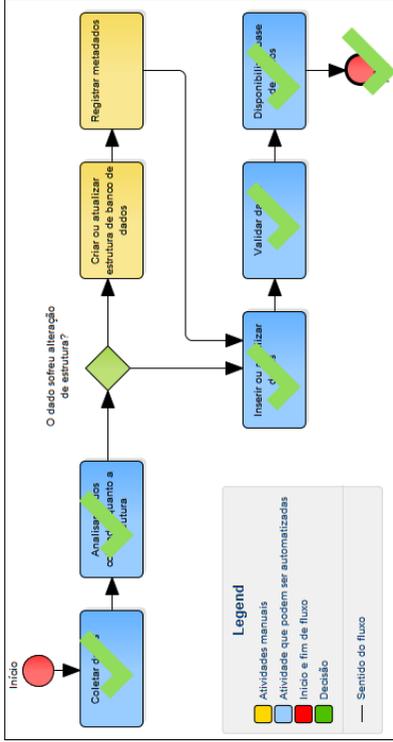


Figura 34 – Atualização de Horário de transporte.  
Fonte: O autor (2018).

## Inclusão de novos dados

Mapeamento dos dados
Processo

**Informações dos dados**

Arquivo (CSV) ...

OcorrênciasAeronauticas.csv

Dado

Número de ocorrências aeronauticas

**Fonte**

Sigla  Nome   
 CIPAA  Centro de Investigação e Prevenção de Acidentes Aeronáuticos

Dados do arquivo

ICAO	QuantidadeDeOcorrências
SBGR	14280504
SBSP	8691526
SBRR	6632917
SBGL	6311763
SBCE	3984792
SBKP	3935156
SBRJ	3732196
SBPA	3103579
SBPF	3076027
SBRE	2968330

**Descrição**

Dados relativos ao quantitativo de ocorrências aeronauticas

Nome da coluna	Tipo de dados	Vincular	Descrição
ICAO	VARCHAR(10)		Código Icao do aer...
QuantidadeDeOcor...	VARCHAR(10)		Número de ocorrên...

Carregar dados

Figura 35 – Inclusão de novos dados.  
 Fonte: O autor (2018).

coluna do novo arquivo com os dados de aeródromos, municípios ou estados. Para isso, basta indicar o relacionamento na coluna Vincular da interface de configuração do arquivo. Após realizar as configurações e clicar no botão Carregar dados, o protótipo de sistema executa os passos propostos no método. Conforme consta na figura 36, como trata-se de um novo dado, o protótipo executa as atividades de criação de estrutura e registro dos metadados.

Quando um novo dado é incluído na base de dados do protótipo de sistema, ele é automaticamente disponibilizado no módulo de dados completos (seção 4.7.1.4), ou seja, o dado fica disponível para manipulação e uso. Na figura 37, é possível visualizar o dado incluído no protótipo e já disponível no módulo de dados completos.

Quando os dados adicionados ao protótipo têm um relacionamento com os dados de aeródromos, é possível utilizá-los para a criação de mapas temáticos específicos. Para isso, é necessário selecionar o novo dado, definir a coluna que se deseja utilizar para a construção do tema e a cor dos pontos, conforme apresentado na figura 38.

Após a seleção e a configuração dos parâmetros, é necessário clicar no botão Executar para que o protótipo de sistema apresente um mapa temático e informações detalhadas (gráfico e *grid* de dados) sobre os dados incluídos (Figura 39).

## Inclusão de novos dados

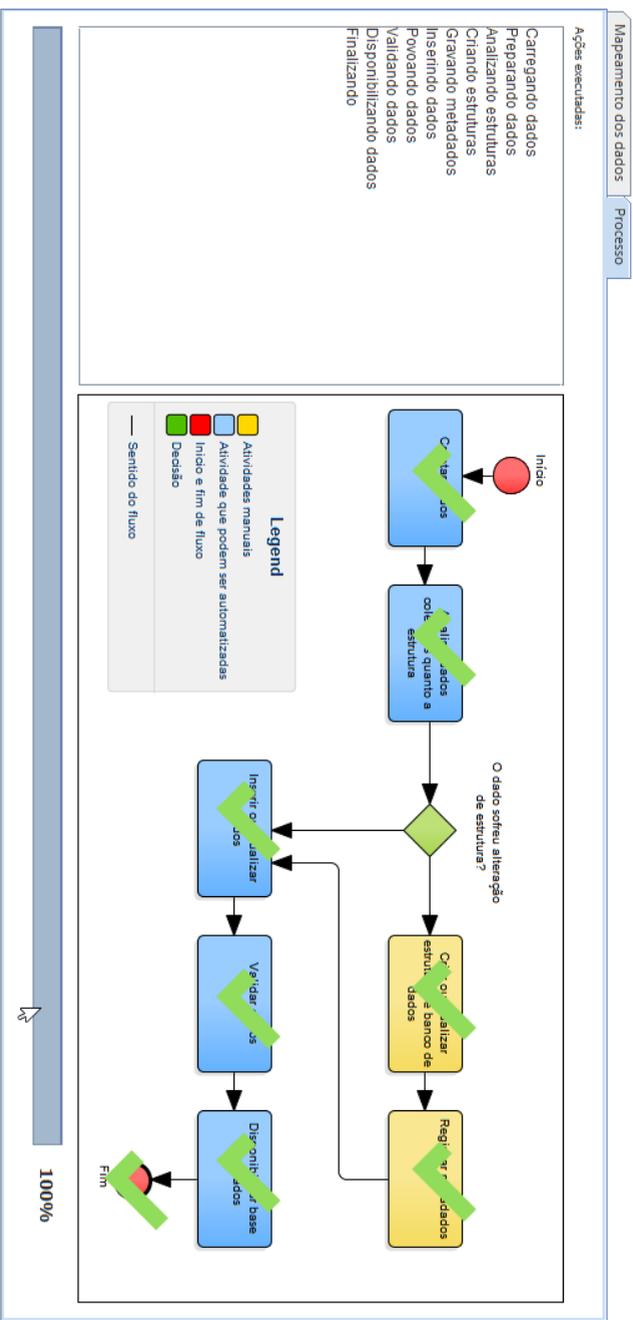


Figura 36 – Processo de vinculação de novos dados.  
 Fonte: O autor (2018).

### Dados completos Número de ocorrências aeronauticas

Carregar

id	nome	codigoicao	aerodromo	municipio	Ocorrências	movimentacao...	taxadecrescime...	taxamediadecl...	posicaooranki...	idAerodromo
51	GUARULHOS GO...	SBGR	Aeroporto Interna...	Guarulhos/SP	14280504	5.671.679,00	-4,94	-2,28	1	51
115	CONGONHAS	S BSP	Aeroporto de São ...	São Paulo/SP	8691526	5.671.679,00	7,27	-2,28	2	115
15	PRESIDENTE JUSC...	SBBR	Aeroporto Interna...	Brasília/DF	6632917	5.671.679,00	-12,74	-2,28	3	15
47	AEROPORTO INTE...	SBGL	Aeroporto Interna...	Rio de Janeiro/RJ	6311763	5.671.679,00	-1,18	-2,28	4	47
23	TANCREDO NEVES	SBCF	Aeroporto Interna...	Confins/MG	3984792	5.671.679,00	-2,67	-2,28	5	23
72	VIRACOPOS	SBKP	Aeroporto Interna...	Campinas/SP	3935156	5.671.679,00	3,68	-2,28	6	72
107	SANTOS DUMONT	SBRJ	Aeroporto do Rio ...	Rio de Janeiro/RJ	3732196	5.671.679,00	0,61	-2,28	7	107
117	DEPUTADO LUÍS E...	SBSV	Aeroporto Interna...	Salvador/BA	3103579	5.671.679,00	-8,5	-2,28	8	117
91	SALGADO FILHO	SBPA	Aeroporto Interna...	Porto Alegre/RS	3076027	5.671.679,00	0,07	-2,28	9	91
106	GUARARAPES GIL...	SBRF	Aeroporto Interna...	Recife/PE	2968330	5.671.679,00	3,71	-2,28	10	106

Total de registros: 10

Exportar CSV

Figura 37 – Visualização dos novos dados.  
Fonte: O autor (2018).

SIDDA - Sistema de Integração e Disponibilização de Dados Aeroportuários X

### Mapa e indicadores temáticos

dados externos

Dados disponíveis

ANAC - VRA

Indicador

Figura 38 – Criação de tema com novos dados de aeródromos.  
Fonte: O autor (2018).

Tabela 3 – Registro da coleta dos dados.

Dado	Nome do arquivo/pasta
IBGE - Indicadores de municípios	20160615_IBGE_IndicadorMunicipio.xls
IBGE - Documentação dos indicadores	20160615_IBGE_MetodologiaIndicador.pdf
ANAC - Aeródromos públicos	20160616_ANAC_AerodromoPublico.xls
ANAC - Aeródromos privados	20160616_ANAC_AerodromoPrivado.xls
ANAC - Voo Regular Ativo	ANAC_VRA*
ANAC - Dados estatísticos	20160616_ANAC_DadoEstatistico.csv
ANAC - Horário de transporte	20160616_ANAC_Hotran.csv
ANAC - Tarifas aéreas domésticas	20160616_ANAC_Tarifa.csv
INFRAERO - Estatísticas	INFRAERO_Estatistica*
SAC/MTPA - Desempenho Operacional	20170813_SAC_DSO.xls
DECEA - NOTRAN	DECEA_NOTRAN*

\* Dados compostos por múltiplos arquivos, portanto, estão armazenados em um diretório específico.

Fonte: O autor (2018).

Tabela 4 – Análise da estrutura dos dados de municípios.

Coluna	Nome para coluna	Tipo de dado
Ano de referência	ano	número inteiro
Código da Unidade da Federação	idUF	número inteiro
Nome da Unidade da Federação	nomeUF	texto
Código do Município	idMunicipio	número inteiro
Nome do Município	nomeMunicipio	texto
Região Metropolitana	nomeRegiaoMetro	texto
Código da Mesorregião	idMesor	número inteiro
Nome da Mesorregião	nomeMesor	texto
Código da Microrregião	idMicror	número inteiro
Nome da Microrregião	nomeMicror	texto
Valor adicionado bruto da Agropecuária, a preços correntes	valorBrutoAgro	número decimal
Valor adicionado bruto da Indústria, a preços correntes	valorBrutoIndustria	número decimal
Valor adicionado bruto dos Serviços, a preços correntes	valorBrutoServico	número decimal
Valor adicionado bruto da Administração, a preços correntes	valorBrutoAdm	número decimal
Valor adicionado bruto Total, a preços correntes	valorBruto	número decimal
Impostos, líquidos de subsídios, sobre produtos, a preços correntes	impostos	número decimal
Produto Interno Bruto, a preços correntes	pib	número decimal
População (Nº de habitantes)	populacao	número decimal
Produto Interno Bruto per capita	pibPerCapita	número decimal

Fonte: Adaptada de IBGE (2017b).

Tabela 5 – Dados compilados de unidade da federação.

Código UF	UF	Nome
12	AC	Acre
27	AL	Alagoas
16	AP	Amapá
13	AM	Amazonas
29	BA	Bahia
23	CE	Ceará
53	DF	Distrito Federal
32	ES	Espírito Santo
52	GO	Goiás
21	MA	Maranhão
51	MT	Mato Grosso
50	MS	Mato Grosso do Sul
31	MG	Minas Gerais
15	PA	Pará
25	PB	Paraíba
41	PR	Paraná
26	PE	Pernambuco
22	PI	Piauí
33	RJ	Rio de Janeiro
24	RN	Rio Grande do Norte
43	RS	Rio Grande do Sul
11	RO	Rondônia
14	RR	Roraima
42	SC	Santa Catarina
35	SP	São Paulo
28	SE	Sergipe
17	TO	Tocantins

Fonte: Adaptada de IBGE (2017b).

Tabela 6 – Amostra dos dados compilados de municípios.

Código UF	Código Microrregião	Código Município	Nome Município
11	11006	1100015	Alta Floresta D'Oeste
11	11003	1100023	Ariquemes
11	11008	1100031	Cabixi
11	11006	1100049	Cacoal
11	11008	1100056	Cerejeiras
11	11008	1100064	Colorado do Oeste
11	11008	1100072	Corumbiara
11	11002	1100080	Costa Marques
11	11006	1100098	Espigão D'Oeste
11	11002	1100106	Guajará-Mirim
11	11004	1100114	Jaru
11	11004	1100122	Ji-Paraná
11	11003	1100130	Machadinho D'Oeste
11	11005	1100148	Nova Brasilândia D'Oeste
11	11004	1100155	Ouro Preto do Oeste
11	11007	1100189	Pimenta Bueno
11	11001	1100205	Porto Velho
11	11004	1100254	Presidente Médici
11	11003	1100262	Rio Crespo

Fonte: Adaptada de IBGE (2017b).

Ano	Código do Município	Valor Bruto Agropecuária	Valor Bruto Indústria	Valor Bruto Serviços	Valor Bruto Administração	Valor Bruto
2010	1100015	69.260	15.890	62.593	93.245	240.988
2010	1100023	73.712	283.673	495.356	343.868	1.196.608
2010	1100031	24.301	3.183	12.495	25.170	65.149
2010	1100049	95.259	180.259	465.156	298.454	1.039.128
2010	1100056	28.976	19.514	80.718	63.018	192.227
2010	1100064	35.818	18.071	54.420	67.462	175.771

Fonte: Adaptada de IBGE (2017b).

Tabela 7 – Dados compilados de indicadores dos municípios (2010).

Tabela 8 – Análise da estrutura dos dados de aeródromos.

Coluna		Nome para coluna	Tipo de dado
Aeródromo	ICAO	icao	texto
	TIPO DE AERODROMO	tipoAerodromo	texto
	NOME	nome	texto
	MUNICÍPIO ATENDIDO	municipio	texto
	UF	uf	texto
Geoespacial	LATITUDE	latitude	número decimal
	LONGITUDE	longitude	número decimal
	ALTITUDE	altitude	número decimal
Operação	OPERAÇÃO	operacao	texto
Pista 1	DESIGNAÇÃO	pista1Designacao	texto
	COMPRIMENTO	pista1Comprimento	número inteiro
	LARGURA	pista1Largura	número inteiro
	RESISTÊNCIA	pista1Resistencia	texto
	SUPERFÍCIE	pista1Superficie	texto
Pista 2	DESIGNAÇÃO	pista2Designacao	texto
	COMPRIMENTO	pista2Comprimento	número inteiro
	LARGURA	pista2Largura	número inteiro
	RESISTÊNCIA	pista2Resistencia	texto
	SUPERFÍCIE	pista2Superficie	texto
Pista 3	DESIGNAÇÃO	pista3Designacao	texto
	COMPRIMENTO	pista3Comprimento	número inteiro
	LARGURA	pista3Largura	número inteiro
	RESISTÊNCIA	pista3Resistencia	texto
	SUPERFÍCIE	pista3Superficie	texto

Fonte: O autor (2018).

Tabela 9 – Amostra dos dados compilados do cadastro de aeródromos.

icao	Nome	Latitude	Longitude	Município
SBAA	CONCEIÇÃO DO ARAGUAIA	-49,3030	-8,3486	Conceição do Araguaia
SBAE	BAURU/AREALVA	-49,0683	-22,1577	Bauru
SBAM	AMAPÁ	-50,8624	2,0727	Amapá
SBAQ	BARTOLOMEU DE GUSMÃO	-48,14028	-21,8044	Araraquara
SBAR	SANTA MARIA	-37,0733	-10,9852	Aracaju

Fonte: Adaptada de ANAC (2017a).

Tabela 10 – Amostra do dados compilados de pistas dos aeródromos.

ICAO	Designador)	Comp.	Larg.	Resistência	Pavimento
SDNW	15/33	800	18	4000kg/0,50 MPa	Aço
SDOX	01/19	500	18	5700kg/0,50MPa	Aço
SITL	03/21	1100	23	5700Kg/0.50MPa	Aço
SSQO	09/27	1300	20	5400kg/0,50MPa	Aço
SJHU	01/19	750	23	5700kg/0,50 MPa	Areia
SNMR	12/30	800	30	5000 kg / 0,5 M	Areia
SNVV	11/29	1000	20	5700 kg / 0,5 M	Areia
SSOS	10/28	1000	40	8/F/C/Y/U	Areia
SNCE	10/28	1000	30	8/F/C/Y/U	Argila
SWAK	6/24	950	30	2500 kg / 0,50	Argila

Fonte: Adaptada de ANAC (2017a).

Tabela 11 – Análise da estrutura dos dados de HOTRAN.

Coluna	Nome para coluna	Tipo de dado
Cód. Empresa	codEmpresa	Texto
Empresa	nomeEmpresa	Texto
Nº VOO	numeroVoo	Texto
Equip.	equipamento	Texto
Dias da semana	diaSemana	Texto
Qtde Assentos	qtdAssento	Número inteiro
Número Hotran	numeroHotran	Texto
Data Solicitação	dataSolicitacao	Data e hora
Data Aprovação	dataAprovacao	Data e hora
Data Vigência	dataVigencia	Data
Natureza Operação	naturezaOperacao	Texto
Nº Etapa	numeroEtapa	Número inteiro
COD. Origem	origemIcao	Texto
ARPT Origem	origemNome	Texto
COD. Destino	destinoIcao	Texto
ARPT Destino	destinoNome	Texto
Horário Partida	partida	Hora
Horário Chegada	chegada	Hora
CODESHARE	codeshare	Texto

Fonte: Adaptada de ANAC (2017c).

Tabela 12 – Análise da estrutura dos dados de movimentação.

Coluna	Nome da coluna	Tipo de dado
Empresa (Sigla)	empresaSigla	Texto
Empresa (Nome)	empresaNome	Texto
Empresa (Nacionalidade)	empresaNacionalidade	Texto
Ano	ano	Número inteiro
Mês	mês	Número inteiro
Aeroporto de origem (Sigla)	aeroportoOrigemSigla	Texto
Aeroporto de origem (Nome)	aeroportoOrigemNome	Texto
Aeroporto de destino (Sigla)	aeroportoDestinoSigla	Texto
Aeroporto de destino (Nome)	aeroportoDestinoNome	Texto
Natureza	natureza	Texto
Grupo de voo	grupoVOO	Texto
Passageiros (Grátis e pago)	passageirosPagos	Número inteiro
Carga (Kg) (Grátis e pago)	cargaKG	Número decimal
Correio (Kg)	correioKg	Número decimal
ASK	ask	Número decimal
RPK	rpk	Número decimal
ATK	atk	Número decimal
RTK	rtk	Número decimal
Combustível (Litros)	combustível	Número decimal
Distância voada (Km)	distânciaKM	Número decimal
Decolagens	decolagens	Número inteiro
Carga (Km) (Grátis e pago)	cargaKM	Número decimal
Correio (Km)	correioKm	Número decimal
Assentos	assentos	Número inteiro
Payload	payload	Número decimal
Horas voadas	horasVoadas	Número decimal

Fonte: Adaptada de ANAC (2017a).

Tabela 13 – Amostra dos dados compilados de companhias aéreas.

Sigla	Nome
2BF	BAIRES FLY
AAL	AMERICAN AIRLINES
ABJ	ABAETÉ
ABW	AirBridgeCargo Airlines
ABX	ABX AIR INC.
ABZ	ATA BRASIL
ACA	AIR CANADA
ACX	ACG AIR CARGO GERMANY GMBH
ADB	ANTONOV AIRLINES
AEA	AIR EUROPA

Fonte: Adaptada de ANAC (2017a).

Tabela 14 – Análise da estrutura dos dados de VRA.

Coluna	Nome para coluna	Tipo de dado
ICAO Empresa Aérea	codEmpresa	Texto
Número Voo	numeroVoo	Número inteiro
Código Autorização (DI)	codAutorizacao	Número inteiro
Código Tipo Linha	tipoLinha	Texto
ICAO Aeródromo Origem	origemICAO	Texto
ICAO Aeródromo Destino	destinoICAO	Texto
Partida Prevista	partidaPrevista	Data e hora
Partida Real	partidaReal	Data e hora
Chegada Prevista	chegadaPrevista	Data e hora
Chegada Real	chegadaReal	Data e hora
Situação Voo	cotuacaoVoo	Texto
Código Justificativa	codigoJustificativa	Texto

Fonte: Adaptada de ANAC (2017g).

Tabela 15 – Exemplo de metadados de tabelas.

Tabela	Nome	Descrição	Fonte
UF	Unidade da Federação	A tabela Unidade da Federação refere-se à estruturação dos dados dos estados brasileiros, contendo informações básicas cadastrais.	IBGE
Município	Município	A tabela de municípios do Brasil contém informações desta entidade, que são uma circunscrição territorial dotada de personalidade jurídica e com certa autonomia administrativa, sendo as menores unidades autônomas da Federação.	IBGE
Indicador	Indicadores IBGE	A tabela Indicadores IBGE dispõe de estimativas do Produto Interno Bruto - PIB dos municípios, a preços correntes e <i>per capita</i> , e do valor adicionado da Agropecuária, Indústria e Serviços, a preços correntes, por meio de um processo descendente de repartição pelos municípios.	IBGE

Fonte: O autor (2018).

Tabela 16 – Comparativo dos dados de horário de transporte.

Natureza da Operação	Planilha ANAC		Base de dados	
	Frequência	Assentos	Frequência	Assentos
Cargueiro Doméstico	1387	30100	1387	30100
Linha Aérea Sistemática	205	645	205	645
Rede Postal	461	8418	461	8418
Doméstica	18400	514534	18400	514534
Cargueiro Internacional	644	0	644	0
Internacional	3362	152835	3362	152835

Fonte: O autor (2018).

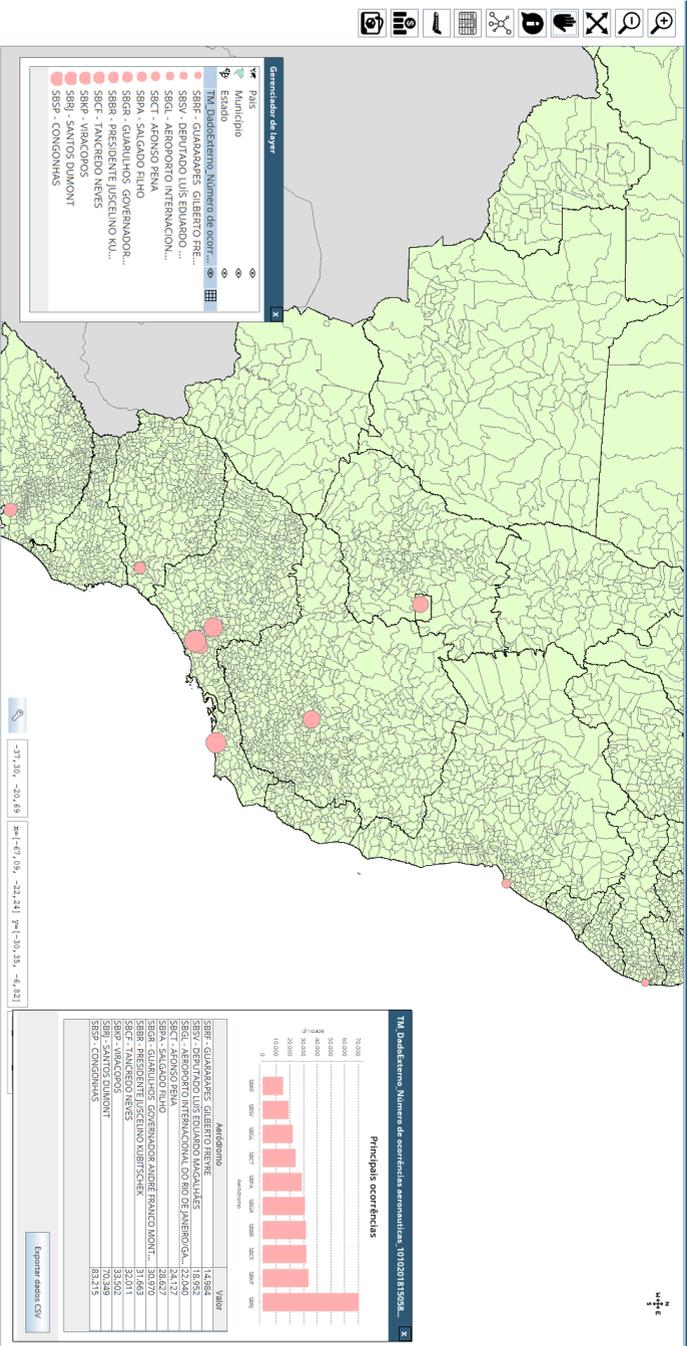


Figura 39 – Tema com novos dados de aeródromos.  
Fonte: O autor (2018).



## 5 EXPERIMENTO DE CARACTERIZAÇÃO DE UM AEROPORTO

Para uma exploração da potencialidade da base de dados e do protótipo de sistema computacional propostos, nesta seção, será apresentado um experimento com o uso dos dados e das funcionalidades. Para isso, foi utilizado o aeroporto Hercílio Luz (SBFL), situado no município de Florianópolis, estado de Santa Catarina. Os gráficos apresentados foram desenvolvidos utilizando-se a ferramenta Microsoft Excel e os dados foram extraídos da base dados construída.

O aeroporto Hercílio Luz fazia parte do grupo de aeroportos indicados pelo Governo Federal disponíveis para novas concessões. Em 16 de março de 2017, o aeroporto foi arrematado em um leilão pela empresa Flughafen Zürich AG, configurando-se como o primeiro caso de gestão 100% estrangeira de um aeroporto nacional.

Segundo dados cadastrais da ANAC (2017a), extraídos do protótipo de sistema, o aeroporto possui as seguintes características:

- **Altitude:** 5 metros (em relação ao nível do mar).
- **Coordenadas:**
  - **Latitude:** -48,5525.
  - **Longitude:** -27,6703.
- **Pista 1:**
  - **Pavimento:** asfalto.
  - **Comprimento:** 2.300 metros.
  - **Largura:** 45 metros.
- **Pista 2:**
  - **Pavimento:** concreto.
  - **Comprimento:** 1.500 metros.
  - **Largura:** 45 metros.

Com os dados de movimentação da ANAC (2017a), foi possível identificar que o aeroporto Hercílio Luz possui registros de movimentação de aeronaves, passageiros, mala postal e carga aérea. Entre o período de janeiro de 2000 e dezembro de 2016, teve um registro de movimentação

expressivo em relação aos demais aeroportos. Ao ranquear o total movimentado dos aeroportos nacionais no período supracitado, o SBFL ocupa a 15<sup>o</sup> posição para movimentação de aeronaves, a 14<sup>o</sup> posição para passageiros, a 19<sup>o</sup> posição para carga aérea e a 16<sup>o</sup> posição para mala postal.

O aeroporto movimentou um total de 478,14 mil aeronaves de janeiro de 2000 a dezembro de 2016 e seu crescimento médio de aeronaves foi de 1,73%. O ano de 2012 foi o que registrou o maior volume de aeronaves, totalizando 37,96 mil (total de operações de pouso e decolagens). Já 2005 indicou apenas 19,19 mil aeronaves, sendo o ano com menor registro. Na figura 40, consta um gráfico com a movimentação histórica de aeronaves do aeroporto Hercílio Luz.

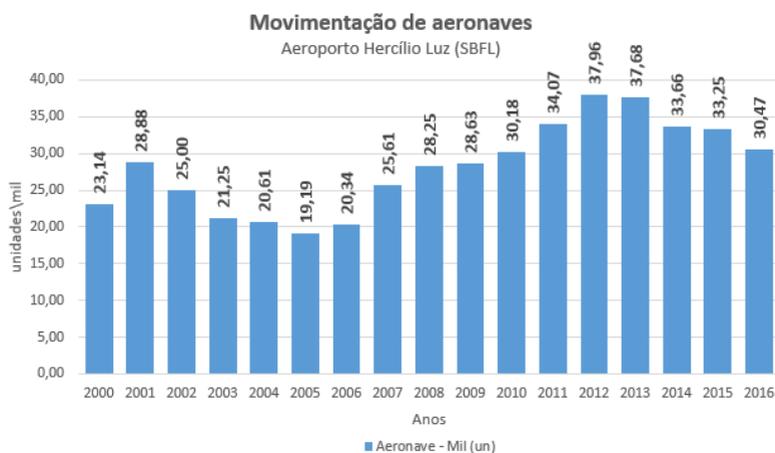


Figura 40 – SBFL - Movimentação de aeronaves.

Fonte: Adaptada de ANAC (2017b).

O aeroporto movimentou um total de 37,99 milhões de passageiros de janeiro de 2000 a dezembro de 2016 e seu crescimento médio de passageiros foi de 8,49%. O ano de 2013 foi o que registrou o maior volume de passageiros, totalizando 3,45 milhões passageiros. Já 2000 indicou apenas 934 mil passageiros, sendo o ano com menor registro. Na figura 41, consta um gráfico da movimentação histórica de passageiros do aeroporto Hercílio Luz.

O aeroporto movimentou um total de 95,25 milhões de toneladas em carga aérea de janeiro de 2000 a dezembro de 2016 e seu crescimento médio de carga aérea foi de -0,25%. O ano de 2006 foi o que registrou o



Figura 41 – SBFL - Movimentação de passageiros.  
Fonte: Adaptada de ANAC (2017b).

maior volume de carga aérea, totalizando 7,47 milhões de toneladas. Já 2002 indicou 4,37 milhões de toneladas, sendo o ano com menor registro de carga aérea. Na figura 42, consta um gráfico da movimentação histórica de carga aérea do aeroporto Hercílio Luz.

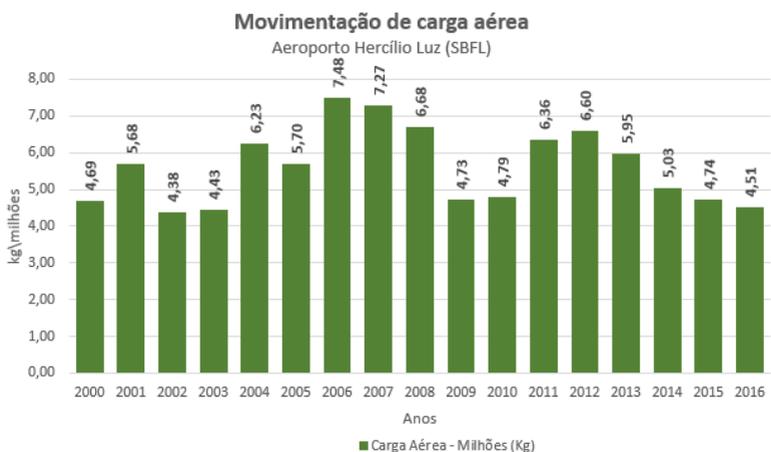


Figura 42 – SBFL - Movimentação de carga aérea.  
Fonte: Adaptada de ANAC (2017b).

O aeroporto movimentou um total de 28,85 bilhões de toneladas em mala postal de janeiro de 2000 a dezembro de 2016 e seu crescimento médio de mala postal foi de -1,95%. O ano de 2001 foi o que registrou o maior volume de mala postal, totalizando 2,84 bilhões de toneladas. Já 2003 indicou 268,79 milhões de toneladas, sendo o ano com menor registro de mala postal. Na figura 43, consta um gráfico da movimentação histórica de mala postal do aeroporto Hercílio Luz.

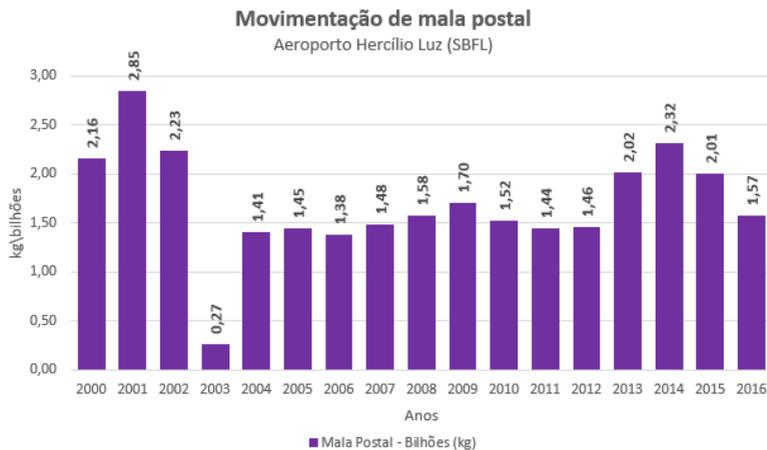


Figura 43 – SBFL - Movimentação de mala postal.  
Fonte: Adaptada de ANAC (2017b).

Considerando os dados históricos, as principais companhias aéreas que operam no aeroporto Hercílio Luz para passageiros são Gol (40% de participação), Latam (33% de participação), Avianca (8% de participação) e Azul (6% de participação). A soma das demais companhias aéreas representa 13% de participação. Tais dados mostram que apenas as companhias Gol e Latam juntas movimentaram um total de 28,40 milhões de passageiros de janeiro de 2000 a dezembro de 2016.

Em relação à série histórica disponível para indicadores socioeconômicos, o município sede do aeroporto (Florianópolis) apresentou, para dados de população, PIB e PIB *per capita*, taxas de crescimento (equação 5.1) e taxas crescimento médio (equação 5.2) positivas. Um destaque é o PIB, com uma taxa de crescimento de 30,02% e taxa média de crescimento de 6,78%, considerando dados de 2010 a 2013. Na tabela 17, é possível visualizar os dados registrados no IBGE para

o município de Florianópolis

$$TxC = \frac{Vp - Vo}{Vo} \times 100 \quad (5.1)$$

Onde:

- TxC: Taxa de crescimento
- Vo: Valor passado
- Vp: Valor presente

$$TxMC = \left( \frac{Vp^{\frac{1}{n}}}{Vo} \right) - 1 \quad (5.2)$$

Onde:

- TxMC: Taxa média de crescimento
- Vo: Valor passado
- Vp: Valor presente
- n: Quantidade de anos

Tabela 17 – Florianópolis - Indicadores socioeconômicos.

Indicadores Socioeconômicos	População (habitantes)	PIB <i>per capita</i> (R\$ 1,00)	PIB (R\$ 1.000)
2010	421.203	26.804,11	11.289.972,00
2011	427.298	29.817,40	12.740.917,00
2012	433.158	31.998,89	13.860.577,00
2013	453.285	32.385,04	14.679.653,00
Tx. Crescimento	7,62%	20,82%	30,02%
Tx. Med. Crescimento	1,85%	4,84%	6,78%

Fonte: IBGE (2017a).

Para identificar cenários futuros quanto à movimentação de passageiros para o aeroporto Hercílio Luz, buscou-se utilizar dados extraídos da base de dados construída e projetá-los em intervalos de cinco anos, de 2020 até 2040. Utilizou-se a ferramenta Microsoft Excel para aplicar a equação 5.3, na qual, para o parâmetro de taxa de crescimento, foi aplicada taxa média de crescimento para a população de

Florianópolis para os anos de 2010 a 2013. Os resultados constam na tabela 17.

Na figura 44, consta um gráfico que gerado a partir dos dados obtidos na projeção da movimentação de passageiros realizada para o aeroporto Hercílio Luz.

$$Vf = Vp \times (1 + c)^a \quad (5.3)$$

Onde:

- Vf: Valor futuro
- Vp: Valor presente
- c: Taxa de crescimento
- a: Período de projeção

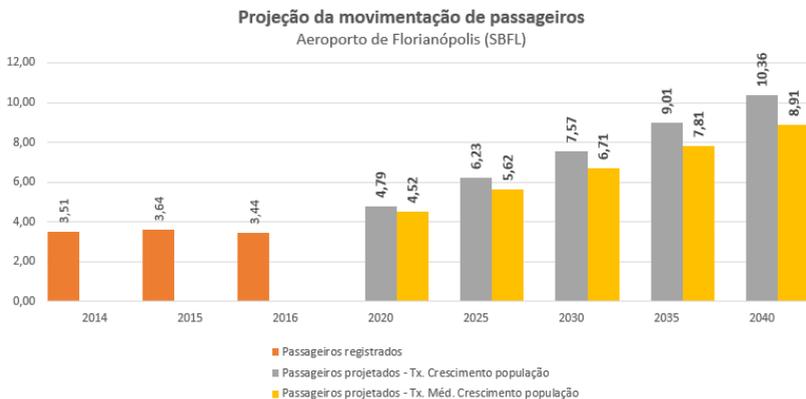


Figura 44 – SBFL - Projeção da movimentação de passageiros.

Fonte: Adaptada de ANAC (2017b) e IBGE (2017a).

Buscando avaliar as principais rotas aéreas do Aeroporto Hercílio Luz, foi realizada uma análise dos dados de HOTRAN. Por meio deles, foi possível identificar que as principais rotas são nos aeroportos de Guarulhos (SBGR) e Congonhas (SBSP), totalizando 121 e 64 voos autorizados semanais, respectivamente. Na figura 45, é possível visualizar as rotas autorizadas para as quais o aeroporto Hercílio Luz é origem ou destino do voo.

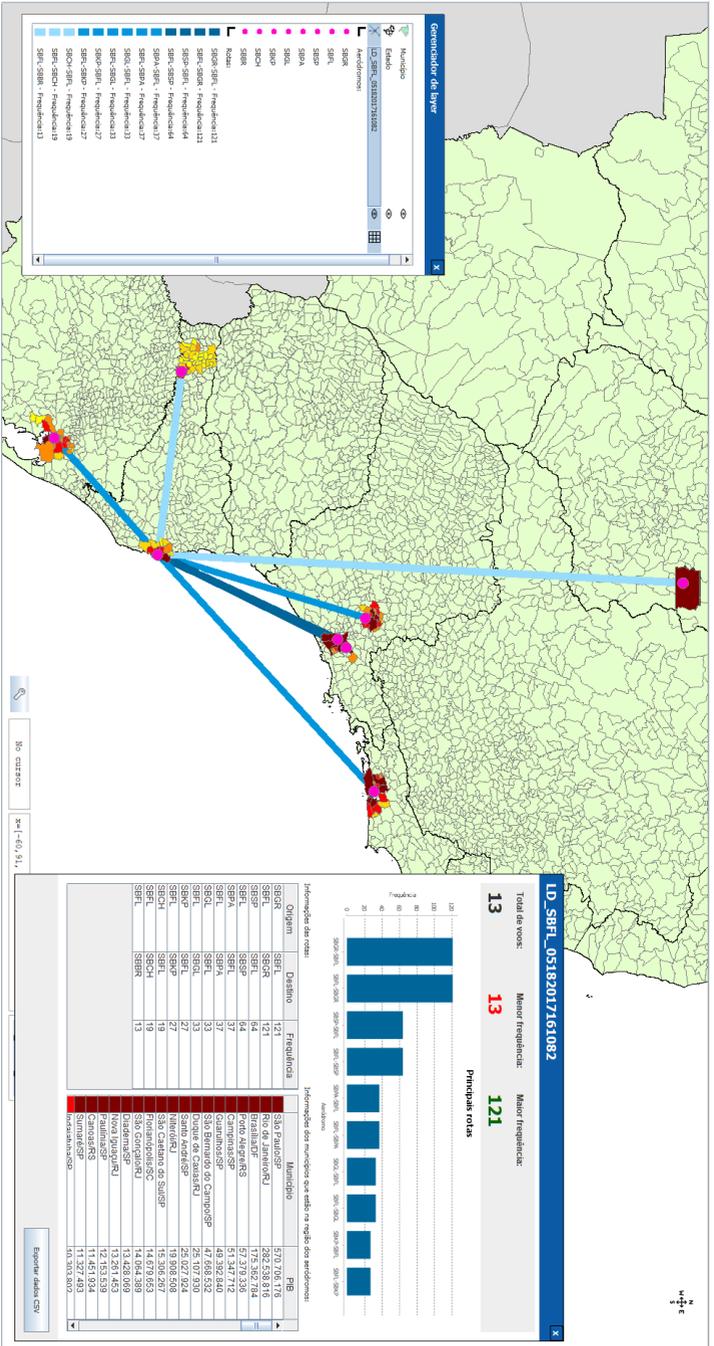


Figura 45 – SBFL - Principais fluxos (linhas de destino).  
 Fonte: Adaptada de ANAC (2017c) e IBGE (2017a).

Considerando os dados de movimentação, no ano de 2016, a rota aérea do aeroporto Hercílio Luiz para o aeroporto de Guarulhos movimentou 521 mil passageiros, enquanto que o sentido contrário movimentou 511 mil. Já a rota aérea do aeroporto Hercílio Luiz para aeroporto de Congonhas movimentou 419 mil passageiros, enquanto que o sentido contrário movimentou 426 mil. Essas rotas juntas (considerando ambos os sentidos), somente no ano de 2016, movimentaram 1,87 milhões de passageiros.

Um dado a ser observado e que pode caracterizar o nível de serviço (quanto à pontualidade com o passageiro) é o percentual de voos cancelados em relação aos voos realizados. Segundo os dados de voo regular ativo, no ano de 2016, o aeroporto Hercílio Luz apresentou um baixo índice de cancelamento – apenas 9% de voos cancelados para 91% de voos realizados.

O aeroporto Hercílio Luz é bastante relevante para a rede aérea nacional, principalmente para a Região Sul do país. Apesar de ocupar a terceira posição na movimentação de passageiros, ficando atrás dos aeroportos Salgado Filho (SBPA) (no Rio Grande do Sul) e Afonso Pena (SBCT) (no Paraná), ao comparar a movimentação desses aeroportos com a população dos municípios sede, o Hercílio Luz apresenta uma relação de movimentação 4,74 vezes a população. Já o aeroporto de Salgado Filho apresenta uma relação 3,45 vezes a população e o Afonso Pena, 2,44 vezes a população. Portanto, correlacionando a movimentação de passageiros e a população dos municípios, pode-se afirmar que o aeroporto Hercílio Luz tem a movimentação mais expressiva da Região Sul do Brasil.

## 6 CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS

Este capítulo visa estabelecer as conclusões a respeito o trabalho desenvolvido no âmbito da dissertação e algumas recomendações deixadas para estudos posteriores relacionados ao tema principal.

### 6.1 CONCLUSÕES

O presente trabalho teve como objetivo a realização de um estudo para o desenvolvimento de uma base de dados integrada, a fim de apoiar tarefas relacionadas ao planejamento da infraestrutura do transporte aéreo brasileiro, focando na utilização de dados públicos. Para atender o objetivo geral, um método para o desenvolvimento de uma base de dados integrada foi proposto, desenvolvido, implementado em computador e teve o seu funcionamento devidamente demonstrado. Para tanto, foram analisados fundamentos teóricos e técnicos, os quais permitiram que se chegasse aos resultados expostos.

Após o estudo da literatura e da varredura de portais governamentais, foram mapeadas as práticas necessárias para que fosse possível estabelecer um método que orientasse a construção e a manutenção de uma base de dados. O processo ETL foi definido como um requisito mínimo para o método, pois trata de extração, transformação e carga de um grande volume de dados – tais tarefas, na prática, são fundamentais para a construção de qualquer base.

O desenvolvimento de uma base de dados para suprir demandas por dados de planejamento do transporte aéreo brasileiro requereu dados de fontes e modos de disponibilização distintos; portanto, verificou-se que a construção da base não se resumiria a uma carga de dados direta em um conjunto de tabelas, como se a base de dados contivesse a mesma estrutura dos dados coletados. Para a elaboração de uma base nesse âmbito, foi necessária não só a identificação dos dados, mas também a realização de análises dos modos de disponibilização deles, de *layout* e volume. Nesse contexto, foram utilizados conceitos de *data mining*, com a finalidade de incluir no método atividades voltadas à análise dos dados que fariam parte da base e, com isso, estabelecer regras para o devido tratamento deles. Verificou-se ainda a importância do registro e do armazenamento de metadados.

Para garantir que as atividades necessárias para a construção de

uma base de dados foram estabelecidas, buscou-se tratar sobre como disponibilizar a base para leigos em sistemas de gerenciamento de banco de dados. Para isso, foi executada a atividade que desenvolveu um protótipo de sistema que permitisse consultar, visualizar os dados e indicadores contidos na base de dados criada. Portanto, em suma, foi estabelecido um conjunto de atividades que compõem um método e que propiciam e orientam a construção de uma base de dados sólida, confiável e passível de ser utilizada por indivíduos sem conhecimentos específicos na área de banco de dados.

O método para a construção de base de dados proposto permite que qualquer pessoa com interesse na temática possa reproduzi-lo. As atividades e os fluxos desenvolvidos foram descritos e vêm acompanhados de questionamentos que auxiliam na identificação quanto ao cumprimento das atividades; isso significa que, ao final de cada uma delas, consta uma pergunta a ser respondida, a fim orientar o indivíduo que está aplicando o método identificar se o objetivo daquela etapa foi atendido ou não.

Durante a aplicação do método, foi confirmado que dados públicos para o planejamento do setor de aviação civil existem e estão disponíveis, conforme levantado durante a revisão da literatura. Assim, nesse processo, foram encontradas seis fontes de dados governamentais que disponibilizam abertamente, em seus portais na internet, dados do setor. Entretanto, não foram identificadas documentações que descrevessem os meios de acesso e o modelo dos dados disponibilizados. No contexto dos estudos de planejamento do setor aeroportuário, é importante que os dados sejam disponibilizados pelos órgãos competentes juntamente com seus metadados; entretanto, muitas vezes, estes são inacessíveis ou inexistentes. Assim, a etapa de coleta trouxe grandes desafios, pois o modo de disponibilização dos dados é variado, ou seja, cada instituição tem um padrão, uma característica contraditória à Lei de Acesso à Informação – cabe ressaltar que tais instituições poderiam utilizar os serviços do Portal de Dados Abertos para a disponibilização de seus dados à sociedade em geral de forma uniforme.

Entre as fontes identificadas, a ANAC pode ser caracterizada como a instituição que disponibiliza a maior quantidade de dados acerca do setor de aviação civil nacional. No entanto, cada conjunto de dados é fornecido de maneira distinta pela Agência, existindo dados acessíveis por meio de *webservices*, arquivos mensais, arquivos consolidados e ainda arquivos de dados compactados.

A SAC/MTPA tem um modo de disponibilização dos dados muito similar ao da ANAC; ambas fornecem os dados por meio de pla-

nilhas. Entretanto, a Secretaria o faz em um formato completo (toda a base de dados em um único arquivo) e ainda oferece visões específicas dos dados por meio de filtros que podem ser aplicados em gráficos e grades. Os arquivos são oriundos de serviços *web* específicos para a informação. Mesmo que os dados da SAC/MTPA sejam, na maioria dos casos, de fonte ANAC, todos são acompanhados do procedimento realizado para o seu tratamento. Tais ações favorecem os usuários quanto ao uso e às necessidades por informações do setor.

O DECEA disponibiliza um grande volume de dados, os quais podem ser vistos como complementares aos disponibilizados pela ANAC. Uma característica que pode ser destacada do DECEA é quanto a seu modo de disponibilização, qual seja, por meio de uma API, que oferece acesso padronizado e direto aos dados. Esse formato adotado oferece aos usuários dos dados grande flexibilidade. Por tratar-se de uma API, o acesso independe de tecnologia, plataforma, ferramenta ou sistema operacional; dessa forma, sua escalabilidade fica diretamente atrelada aos recursos computacionais dos usuários dos dados.

O IBGE é considerado o maior acervo de informações do Brasil, mas é também o órgão que disponibiliza dados do modo mais precário (tecnologicamente) entre as fontes identificadas. Seus dados estão espalhados em seu portal e muitas vezes disponíveis em sistemas FTP, cujas informações encontram-se pulverizadas em pastas organizadas por ano. Esse modo de disponibilização dificulta tanto a busca por informações quanto a realização de procedimentos de coleta, pois não existe um padrão de navegação entre as pastas.

Já os dados estatísticos disponibilizados pela INFRAERO estão estruturados em um formato que dificulta sua utilização; cada linha representa um tipo de operação de um aeroporto, não existindo detalhamento necessário para facilitar o processamento. O conjunto de dados de aeroportos disponibilizados (somente os que estão sob a sua gestão) já está contemplado nos dados da ANAC, porém, na Agência, estão em um formato que facilita seu processamento e sua análise.

Com o estabelecimento das fontes de dados e a identificação de quais dados estavam disponíveis, seus endereços de acesso (URL) e suas particularidades (*layout* e modo de acesso), iniciou-se o desenvolvimento de rotinas de *textitsoftware* que auxiliassem nesse processo de coleta. Assim, a partir de modos distintos de acesso para cada grupo de dados, estes foram coletados e analisados quanto à sua estrutura e ao seu conteúdo.

A partir desse contexto e conforme o método proposto, partiu-se para a efetiva construção da base de dados integrada, foco deste es-

tudo. Durante o seu desenvolvimento, identificou-se a necessidade de manipulação dos dados disponíveis, a fim de convertê-los para um formato compatível com a estrutura de armazenamento proposta. Nesse processo, foi identificada a possibilidade de geração de novos conhecimentos (novas informações) sobre o setor, o que demanda que os dados devem estar devidamente atualizados e disponíveis.

A base de dados construída contempla uma visão de armazenamento de informações voltadas ao negócio de aviação civil e outra para informações de controle do protótipo de sistema. Na visão de negócio, muitos dos dados podem ser atualizados mensalmente pelos usuários do protótipo de sistema, tais como cadastro de aeródromos (3 mil registros), VRA (100 mil registros), desempenho operacional (268 mil registros) e dados estáticos (780 mil registros). Já os dados de HOTRAN disponibilizam uma média de 5 mil registros por dia. Com isso, estima-se – com base no volume inserido em procedimentos de atualização de dados – que, na base de dados, possivelmente será inserida ou atualizada uma média de 2 milhões de registros mensalmente.

Atrelada à base de dados, têm-se o SIDDA que é um protótipo de sistema com características de ferramentas SIG, que auxilia na visualização dos dados e indicadores acerca do setor aeroportuário. Este, apoia tomada de decisão e oferece serenidade para os órgãos e iniciativas de uso de base de dados integradas. O mesmo mostrou-se eficiente quanto à disponibilização e à apresentação dos dados e indicadores de desempenho constantes na base de dados integrada. Pois nele, os dados são apresentados em mapas, gráficos e grade de dados. Os módulos de atualização e inclusão de dados seguem o método proposto para garantir a perenidade da base, e sua disponibilidade demonstra que o método pode executado de forma manual ou automatizada.

A tecnologia elaborada para desenho de mapas (utilizando a biblioteca GeoTools) para o desenvolvimento do protótipo de sistema com características de SIG, bem como para a manipulação de dados espaciais, dispõe de recursos que são passíveis de evolução quanto aos modos de apresentação dos resultados. Foram disponibilizadas aos usuários funcionalidades básicas de um SIG, tais como criação de mapas e indicadores temáticos para formas geográficas do tipo área (por exemplo, municípios), linha (por exemplo, rotas entre aeródromos) e pontos (por exemplo, aeródromos).

O módulo de dados completos cumpre seu objetivo quanto à disponibilização de dados no formato em que estes se encontram na base. Nele, constam todos os dados incorporados à base de dados, bem como visões e dimensões específicas dos dados. Em seu contexto, está

a funcionalidade de exportação dos dados para uso em outros sistemas.

No âmbito do módulo de principais rotas entre aeródromos, seu resultado, qual seja, a apresentação dos dados em formato de linhas de desejo, agregando a temática do PIB dos municípios próximos ao aeroporto, é importante e de grande serventia para a identificação de demanda e possibilidades de voos entre os aeródromos. Como possibilidades de melhoria, indica-se que o módulo poderia permitir a utilização de dados de movimentação ou VRA, possibilitando não só a visualização do que está planejado de voo, mas também do executado. Na parte de municípios, o protótipo de sistema poderia ainda permitir a utilização de outros indicadores socioeconômicos. Essas possibilidades dariam ao usuário um grande potencial de uso das informações.

O módulo de temáticas facilita a visualização dos dados de movimentação por meio de um mapa e poderia, em uma evolução, incorporar estilos diferenciados para os elementos geográficos, bem como a disponibilização de outros filtros para a seleção do dado. Para os dados disponíveis na funcionalidade, é passível a inclusão de filtros específicos para os conjuntos de dados, oferecendo flexibilidade quanto à criação de mapas.

Conclui-se que, com a aplicação do método, foi possível alcançar as metas propostas, bem como justificar a sua utilização em eventuais iniciativas de construção de bases de dados voltadas ao planejamento de transportes. Considera-se que o método mostrou-se consistente, tendo como resultados uma base de dados acerca do setor de aviação civil nacional e um protótipo de sistema para a visualização e a análise dos dados.

A utilização do protótipo de sistema desenvolvido a nível institucional é factível e representaria um grande facilitador do que diz respeito ao acesso a dados relevantes para projetos em níveis de planejamento e também de operação de aeroportos. É notório, no entanto, que para viabilizar essa ação, é necessário um programa de dimensionamento de infraestrutura de *software*, planos de capacitação, bem como a consolidação do protótipo transformando o mesmo num produto. Um plano de capacitação também permitiria maior divulgação dos dados disponíveis, a identificação de novas necessidades e o nivelamento quanto ao uso da solução.

Este trabalho pretende contribuir tanto para a academia quanto para iniciativas governamentais, as quais estejam baseadas na construção, na manutenção e na atualização de bases de dados para o planejamento de transportes. Com o uso do método proposto, será possível não apenas desenvolver novas bases, mas também reaproveitar bases de

dados existentes, implicando também no desenvolvimento e no aprimoramento de documentação para mantê-las atualizadas. Além disso, um sistema computacional como o proposto permitirá que qualquer pessoa tenha acesso às bases de dados e as mantenham devidamente atualizadas.

## 6.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Recomenda-se às instituições fornecedoras de dados que as mesmas busquem, cada vez mais, a divulgação dos dados em formato aberto e de forma sistemática. Em conjunto aos dados é recomendado a disponibilização da documentação detalhada acerca de seus formatos e conteúdo. Dessa maneira, a coleta, a utilização e a ampliação de aplicações voltadas ao seu uso seriam facilitadas. Além disso, as instituições além de fornecerem informações à sociedade, estariam atendendo à Lei de Acesso à Informação.

Em relação ao protótipo, seus módulos e funcionalidades podem ser evoluídos e novos módulos podem ser incorporados ao escopo. Ainda, sua potencialidade pode ser expandida se disponibilizada em uma plataforma *web*, pois o tempo para disponibilização de novos recursos e dados seria reduzido. Já a base de dados poderia ser subsídio para a construção de um *Big Data* georreferenciado, que permitiria identificar novos padrões nos dados e, quando associado a uma solução *web*, poderia garantir novas formas de apresentação deles.

Por fim, propõe-se a realização de uma pesquisa acerca dos aspectos de Interconexão; segurança; meios de acesso; organização e intercâmbio de informações; e áreas de integração para Governo Eletrônico, presentes na documentação de Padrões de Interoperabilidade de Governo Eletrônico (e-PING) buscando a realização da aderência entre o método proposto e as práticas governamentais existentes no e-PING.

## REFERÊNCIAS

- ABREU, F. E. L. V. **Análise Sistêmica do Setor Aéreo Brasileiro: Propostas para o Planejamento do Setor.** Dissertação (Dissertação de Mestrado) — Departamento de Engenharia Civil e Ambiental da Faculdade de Tecnologia da Universidade de Brasília, 2008.
- AKKAOUI, Z. E. et al. **Data Warehousing and Knowledge Discovery: 14th International Conference, DaWaK 2012, Vienna, Austria, September 3-6, 2012. Proceedings.** 1. ed. [S.l.]: Springer-Verlag Berlin Heidelberg, 2012. (Lecture Notes in Computer Science 7448).
- ANAC. Aeroportos. Acessado em: 18/12/2017. 2017. Disponível em: <<https://goo.gl/A6iIW1>>.
- ANAC. Base de dados estatísticos do transporte aéreo. Acessado em: 18/12/2017. 2017. Disponível em: <<https://goo.gl/1NQc7G>>.
- ANAC. Hotran. Acessado em: 18/12/2017. 2017. Disponível em: <<https://goo.gl/6W7pTM>>.
- ANAC. **Manual para Membros Consultivos do HOTRAN.** [S.l.], 2017. Acessado em: 02/11/2017. Disponível em: <<https://goo.gl/zv1y3W>>.
- ANAC. Microdados de tarifas aéreas comercializadas. Acessado em: 18/12/2017. 2017. Disponível em: <<https://goo.gl/8uZLYJ>>.
- ANAC. Página institucional. Acessado em: 23/10/2017. 2017. Disponível em: <<https://goo.gl/ObzpQT>>.
- ANAC. Voo regular ativo. Acessado em: 15/12/2017. 2017. Disponível em: <<https://goo.gl/bkzkXA>>.
- ARFAOUI, N.; AKAICHI, J. Generating data mart schema from olap requirements. In: **2013 3rd International Conference on Information Technology and e-Services (ICITeS).** [S.l.: s.n.], 2013. p. 1–6.
- ASHFORD, N. J.; MUMAYIZ, S.; WRIGHT, P. H. **Airport Engineering Planning, Design, and Development of 21st**

**Century Airports.** Fourth edition. [S.l.]: John Wiley & Sons, Inc, 2011.

BARBIERI, C. **BI2: Business Intelligence.** [S.l.: s.n.], 2011.

BRASIL. **LEI Nº 7.565.** 12 1986. Disponível em: <<https://goo.gl/z3gZYE>>.

BRASIL. **Decreto Federal nº6.666 de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo Federal, a Infraestrutura Nacional de Dados Espaciais - INDE.** 11 2008. Diário Oficial da União, Poder Executivo, Brasília, DF.

CASTERS, M.; BOUMAN, R.; DONGEN, J. van. **Pentaho Kettle Solutions: Building Open Source ETL Solutions with Pentaho Data Integration.** [S.l.]: Wiley, 2010.

CHAPMAN, P. et al. **CRISP-DM 1.0 Step-by-step data mining guide.** [S.l.], Agosto 2000. Disponível em: <<http://www.crisp-dm.org/CRISPWP-0800.pdf>>.

CONCAR. Plano de ação para implantação da infraestrutura nacional de dados espaciais. Acessado em: 05/09/2017. 2009. Disponível em: <<https://goo.gl/JjYZ7g>>.

DAMA. **DAMA - Guia para o corpo de conhecimento em gerenciamento de dados.** [S.l.]: Technics Publications, LLC, 2012.

DATE, C. J. **An introduction to database systems.** 8ed.. ed. [S.l.]: Addison Wesley Longman, 2003.

DECEA. Página institucional. Acessado em: 01/09/2017. 2017. Disponível em: <<https://goo.gl/wcbzs8>>.

DUPOR, S.; JOVANOVIĆ, V. An approach to conceptual modelling of etl processes. In: **2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO).** [S.l.: s.n.], 2014. p. 1485–1490.

ELMASRI, R.; NAVATHE, S. B. **Sistemas de Banco de Dados.** 4. ed. [S.l.]: Addison Wesley, 2005.

FDT. **Guidebook for Airport Master Planning.** April 2010. Acessado em: 09/06/2016. Disponível em: <<https://goo.gl/sSv4iG>>.

FILHO, M. A. N. de A. **Análise do Processo de Planejamento dos Transportes como Contribuição para a Mobilidade Urbana Sustentável**. Tese (Doutorado) — Universidade de São Paulo, 2012.

GIRARDI, R. A. D. **Framework para coordenação e mediação de Web Services modelados como Learning Objects para ambientes de aprendizado na Web**. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro, 2004.

HOLANDA, L. M. C. de et al. Data webhouse: A evolução do data warehouse para web e suas contribuições para o aperfeiçoamento do relacionamento com clientes. **SEGeT - Simpósio de Excelência em Gestão e Tecnologia**, 2006.

HORONJEFF, R. et al. **Planning and Design of Airports**. Fifth edition. [S.l.]: McGraw-Hill Companies, Inc., 2010.

IBGE. Ibge cidades. Acessado em: 02/10/2017. 2017. Disponível em: <<https://goo.gl/mWCEgQ>>.

IBGE. Portal institucional. Acessado em: 01/08/2017. 2017. Disponível em: <<https://goo.gl/FsidWi>>.

ICAO. **LOCATION INDICATORS - Doc 7910/119**. [S.l.], 2006.

INFRAERO. Estatísticas. Acessado em: 25/09/2017. 2017. Disponível em: <<https://goo.gl/FP6Ujj>>.

INFRAERO. Página institucional. Acessado em: 25/09/2017. 2017. Disponível em: <<https://goo.gl/5eMVf5>>.

INMON, W. H. **Building the Data Warehouse**. 4th ed. ed. [S.l.]: Wiley, 2005.

KAZDA, A.; CAVES, R. E. **Airport Design and Operation**. Second edition. [S.l.]: Elsevier, 2007.

KIMBALL, R. **The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses**. [S.l.]: Wiley, 1996.

KIMBALL, R.; CASERTA, J. **The data warehouse ETL toolkit: practical techniques for extracting, cleaning, conforming, and delivering data**. [S.l.]: Wiley, 2004.

KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling**. 2. ed. [S.l.]: Wiley, 2002.

KIMBALL'S, R. **The data warehouse toolkit: practical techniques for building dimensional data warehouses**. [S.l.: s.n.], 1996.

MALINOWSKI, E.; ZIMANYI, E. **Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications (Data-Centric Systems and Applications)**. 1st ed. 2008. corr. 2nd printing. ed. [S.l.]: Springer, 2011.

MCKINSEY&COMPANY. **Estudo do Setor de Transporte Aéreo do Brasil**. [S.l.], 2010.

MD. Norma da especificação técnica para estruturação de dados geoespaciais vetoriais de defesa da força terrestre. Acessado em: 10/09/2017. 2016. Disponível em: <<https://goo.gl/yHRufc>>.

MDIC. Página institucional. Acessado em: 01/10/2017. 2017. Disponível em: <<https://goo.gl/n3Udf2>>.

MDIC. Sistema de análise das informações de comércio exterior (alice web). Acessado em: 02/10/2017. 2017. Disponível em: <<https://goo.gl/xpGW2c>>.

MORAES, N. G. de. **Avaliação das Tendências da Demanda de Energia no Setor de Transportes no Brasil**. Tese (Doutorado), Rio de Janeiro/RJ, 2005.

MPOG; CONCAR. Plano de ação para implantação da infraestrutura nacional de dados espaciais do brasil (inde). Acessado em: 05/09/2017. 2010. Disponível em: <<https://goo.gl/yHRufc>>.

MPOG; CONCAR. Perfil de metadados geoespaciais do brasil (perfil mgb). Acessado em: 06/09/2017. 2011. Disponível em: <<https://goo.gl/FExBva>>.

OLIVEIRA, B.; BELO, O. Etl process modeling using bpmn meta-models. In: **2013 8th Iberian Conference on Information Systems and Technologies (CISTI)**. [S.l.: s.n.], 2013. p. 1–6. ISSN 2166-0727.

PADULA, R. **Transportes: Fundamentos e propostas para o Brasil**. [S.l.]: Pensar o Brasil - Construir o Futuro da Nação, 2008.

PAIXÃO, S. K. S.; NICHOLS, S.; COLEMAN, D. Towards a spatial data infrastructure: Brazilian initiatives. **Revista Brasileira de Cartografia**, n. 60/2, 08 2008.

PREMA, A.; CLARA, V. T.; PETHALAKSHMI, A. A comparative analysis of decision making methodologies in sales data mart. In: **2014 World Congress on Computing and Communication Technologies**. [S.l.: s.n.], 2014. p. 108–113.

QUEIROZ, M. P. **Definição de Diretrizes para Autorizações de Voos para as Empresas Aéreas**. Tese (Doutorado) — Universidade de Brasília, 2014.

SAC. Relatório de gestão do exercício de 2014. Acessado em: 07/06/2016. 2014. Disponível em: <<https://goo.gl/xSoVh2>>.

SAC. Pesquisa de satisfação. Acessado em: 02/09/2017. 2017. Disponível em: <<https://goo.gl/zcFy3J>>.

SAC. Página institucional. Acessado em: 02/09/2017. 2017. Disponível em: <<https://goo.gl/EjBb4N>>.

SAC. Sistema hórus. Acessado em: 02/09/2017. 2017. Disponível em: <<https://goo.gl/M49pca>>.

SHAW, S. et al. **Practical Hive: A Guide to Hadoop's Data Warehouse System**. 1. ed. [S.l.]: Apress, 2016.

SIMÕES, A. F. **O Transporte Aéreo Brasileiro no Contexto de Mudanças Climáticas Globais: Emissões de CO2 e Alternativas de Mitigação**. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2003.

TADEU, H. F. B. **Cenários de Longo Prazo para o Setor de Transportes e Consumo de Combustíveis**. Tese (Doutorado) — Universidade Católica de Minas Gerais, Belo Horizonte, 2010.

TCU. **5 Motivos para a abertura de dados na Administração Pública**. [S.l.], 2015. Acessado em 31/05/2016. Disponível em: <<https://goo.gl/Hnd5jc>>.

TEOREY, T. et al. **Projeto e Modelagem de Banco de Dados**. Informática. [S.l.]: CAMPUS - GRUPO ELSEVIER, 2013. (Em Português do Brasil).

VAISMAN, A.; ZIMANYI, E. **Data Warehouse Systems: Design and Implementation**. 1. ed. [S.l.]: Springer-Verlag Berlin Heidelberg, 2014. (Data-Centric Systems and Applications).

W3C. Manual dos dados abertos: governo. Acessado em 29/09/2017. 2017. Disponível em: <<https://goo.gl/4uIy8F>>.

WESTERMAN, P. **Data Warehousing: Using the Wal-Mart Model**. 1st. ed. [S.l.]: Morgan Kaufmann, 2000. (The Morgan Kaufmann Series in Data Management Systems).

WOJCIECHOWSKI, J. **JCARBON - SOFTWARE NA WEB COM DATA MINING PARA ESTIMATIVAS DE VOLUME, BIOMASSA E CARBONO EM FLORESTAS**. Tese (Doutorado) — Universidade Federal do Paraná, 2015.

ZHAO, X.; HUANG, Z. A formal framework for reasoning on metadata based on cwm. In: \_\_\_\_\_. **Conceptual Modeling - ER 2006: 25th International Conference on Conceptual Modeling, Tucson, AZ, USA, November 6-9, 2006. Proceedings**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.