



INSTITUTO TECNOLÓGICO DE AERONÁUTICA

Estudos para Aviação do Hoje e do Amanhã

TED n. 11525720240005-003882/2024

PRODUTO 1

## AIRDATA – Artificial Intelligence for Aviation Data Analysis



airdata



## Histórico de versões:

| <i>Versão</i> | <i>Data</i>  | <i>Responsável</i>                | <i>Descrição da Alteração</i>                      |
|---------------|--------------|-----------------------------------|--|
| 1.0           | 9 ago. 2025  | Prof. Dr. Marcelo Xavier Guterres | Versão inicial do produto I                        |
| 2.0           | 25 ago. 2025 | Prof. Dr. Marcelo Xavier Guterres | Versão revisada com os comentários da ANAC e DECEA |
|               |              |                                   |  |





**Coordenação Geral**

Prof. Dr. Cláudio Jorge Pinto Alves  
claudioj@ita.br

**Gerente da Etapa**

Prof. Dr. Marcelo Xavier Guterres  
guterres@ita.br



### **Equipe ITA**

Prof. Dr. Flávio Mendes Neto

Prof. Dr. Dimas Betioli Ribeiro

Prof. Dr. Claudio Sonaglio Albano

Prof. Dr. João Basílio Tarelho Szenczuk

Msc. Guilherme Trindade Tolentino Bernardo

Msc. Jean Phelipe de Oliveira Lima

Vitor Lucas Kohls Correa

Heyder Falheiro de Almeida





# Produto 1

Meta 1 | Etapa 6: AirData

Relatório de Análise e Mapeamento das Bases de Dados



## Sumário Executivo

O presente Produto faz parte da relação de entregas da Etapa 6 da Meta 1 - do TED - Termo de Execução Descentralizada n. 1525720240005-003882/2024, firmado entre a Secretaria de Aviação Civil, cujo número de Processo é 50020.008564/2024-14. Tal TED foi decorrente de estruturação entre a SAC – Secretaria de Aviação Civil e o ITA – Instituto Tecnológico de Aeronáutica, com foco em “*Estudos para Aviação de Hoje e do Amanhã*”. O ITA respondeu à demanda da SAC e o TED citado foi estruturado em 02 (duas) Metas com 16 (dezesesseis) Etapas. O citado TED foi firmado no dia 20/12/2024.

### *Dados Referenciais:*

- TED n. 1525720240005-003882/2024
- Processo n. 50020.008564/2024-14
- Etapa 6 da Meta 1
- Produto I

Para tanto, o presente relatório documenta o primeiro Produto do **AirData**, cujo objetivo central foi mapear, avaliar e propor mecanismos de integração para as principais bases de dados relacionadas ao transporte aéreo brasileiro, visando à construção de uma Base de Conhecimento Semântica que suporte consultas automatizadas por modelos de inteligência artificial.

A etapa incluiu o levantamento e a caracterização das principais fontes de informação, abrangendo bases de operações de voo como SIROS, Dados Estatísticos da ANAC, BIMTRA, TATIC, Síntese Radar, FlightRadar24 e OpenSky Network; dados meteorológicos provenientes do REDEMET, INMET, ERA5 e NCEP/NCAR; informações aeronáuticas oriundas do AISWEB, registros de ocorrências e dados relativos à transição AIS para o AIM; além de bases normativas compostas por RBACs, RBHAs, Instruções Suplementares, Resoluções e Portarias da ANAC e legislação complementar, bem como políticas de dados abertos.

Foram também definidos critérios de qualidade e maturidade para essas bases, com ênfase em completude, atualidade, coerência e interoperabilidade. Como resultado,



propôs-se um esboço de ontologia aeronáutica unificada e fluxos automatizados para coleta, limpeza, enriquecimento semântico e versionamento incremental dos dados, estabelecendo as bases técnicas para a fase seguinte do projeto **AirData**.



## Conteúdo

|   |           |
|---|-----------|
| <b>Lista de Siglas</b>  | <b>6</b>  |
| <b>1 Contextualização</b>                                       | <b>12</b> |
| <b>2 Bases de Operações de Voo</b>                              | <b>15</b> |
| 2.1 Sistema de Registro de Operações (SIROS)                    | 15        |
| 2.2 Dados Estatísticos - Microdados                             | 19        |
| 2.3 Banco de Informações do Movimento de Tráfego Aéreo (BIMTRA) | 28        |
| 2.4 Total Air Traffic Information Control (TATIC)               | 31        |
| 2.5 Síntese RADAR   | 34        |
| 2.6 FlighRadar24 (FR24)   | 36        |
| 2.7 OpenSky Network (OSN)                                       | 39        |
| 2.8 Desafios de Integração entre Bases de Dados                 | 43        |
| 2.9 Considerações Finais  | 49        |
| <b>3 Bases de Dados Meteorológicos</b>                          | <b>50</b> |
| 3.1 REDEMET   | 50        |
| 3.1.1 Meteorological Aerodrome Report (METAR)                   | 51        |
| 3.1.2 Terminal Aerodrome Forecast (TAF)                         | 52        |
| 3.1.3 Sistema de Tempo Severo Convectivo (STSC)                 | 54        |
| 3.1.4 SIGMET  | 55        |
| 3.1.5 GAMET   | 56        |



|          |  |           |
|----------|--|-----------|
| 3.1.6    | RADAR . . . . .  | 59        |
| 3.2      | Instituto Nacional de Meteorologia (INMET) . . . . .                                     | 62        |
| 3.3      | European Analysis 5th Generation (ERA5) . . . . .  | 65        |
| 3.4      | NCEP-NCAR Reanalysis I e NCEP/DOE Reanalysis II (NCEP) . . . . .                         | 67        |
| 3.4.1    | NCEP-NCAR Reanalysis I . . . . .   | 68        |
| 3.4.2    | NCEP/DOE Reanalysis II . . . . .   | 69        |
| 3.5      | Considerações Finais . . . . .   | 70        |
| <b>4</b> | <b>Bases de Informações Aeronáuticas</b>   | <b>74</b> |
| 4.1      | Serviço de Informação Aeronáutica (Serviço de Informação Aeronáutica (AISWEB)) . . . . . | 74        |
| 4.1.1    | Transição de AIS para AIM com base no AIRM . . . . .                                     | 76        |
| 4.2      | Banco de Dados de Ocorrências Aeronáuticas . . . . .                                     | 77        |
| 4.3      | Considerações Finais . . . . .   | 79        |
| <b>5</b> | <b>Bases Normativas</b>  | <b>80</b> |
| 5.1      | Considerações Finais . . . . .   | 82        |
| <b>6</b> | <b>Dados Abertos</b>   | <b>84</b> |
| 6.1      | Conceitos . . . . .  | 84        |
| 6.2      | Premissas e Questões Técnicas . . . . .  | 84        |
| 6.3      | Considerações Finais . . . . .   | 86        |
| <b>7</b> | <b>Propostas de Integração entre Bases de Dados</b>                                      | <b>87</b> |
| 7.1      | Desafios das Bases de Dados Analíticos . . . . .   | 88        |
| 7.2      | Desafios da Base de Conhecimento Semântica . . . . .                                     | 90        |



|          |   |            |
|----------|---|------------|
| 7.3      | Ontologias para o <b>AirData</b> . . . . .          | 91         |
| 7.4      | Grafos de Conhecimento . . . . .                    | 94         |
| 7.5      | Base de Conhecimento Semântica . . . . .            | 98         |
| 7.6      | Convergência e Geração de Respostas . . . . .       | 101        |
| 7.7      | Benefícios Técnicos e Ganhos Operacionais . . . . . | 103        |
| 7.8      | Considerações Finais . . . . .                      | 104        |
| <b>8</b> | <b>Considerações Finais</b>                         | <b>105</b> |



## Lista de Siglas

**ADS-B** *Automatic Dependent Surveillance-Broadcast*

**AIM** Gerenciamento da Informação Aeronáutica

**AIP** Publicação de Informação Aeronáutica

**AIRAC** *Aeronautical Information Regulation And Control*

**AIRM** *Aeronautical Information Reference Model*

**AIS** Serviço de Informação Aeronáutica

**AISWEB** Serviço de Informação Aeronáutica

**AIXM** *Aeronautical Information Exchange Model*

**AMXM** *Airspace Management Exchange Model*

**ANAC** Agência Nacional de Aviação Civil

**API** *Application Programming Interface*

**ATS** *Air Traffic Services*

**BIMTRA** Banco de Informações do Movimento de Tráfego Aéreo

**CENIPA** Centro de Investigação e Prevenção de Acidentes Aeronáuticos

**CNJ** Conselho Nacional de Justiça

**CSV** *Comma-Separated Values*

**DECEA** Departamento de Controle do Espaço Aéreo

**DOE** *Department of Energy*

**ECMWF** *European Centre for Medium-Range Weather Forecasts*

**ERA5** *ECMWF Reanalysis v5*

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



**FIR** *Flight Information Region*

**FIXM** *Flight Information Exchange Model*

**FR24** FlightRadar24

**GRIB** *General Regularly-distributed Information in Binary form*

**IATA** *International Air Transport Association*

**ICA** Instruções do Comando da Aeronáutica

**ICAO** *International Civil Aviation Organization*

**INMET** Instituto Nacional de Meteorologia

**IS** Instruções Suplementares

**IWXXM** *ICAO Meteorological Information Exchange Model*

**JSON** *JavaScript Object Notation*

**LGPD** Lei Geral de Proteção de Dados Pessoais

**LLM** *Large Language Models*

**METAR** *Meteorological Aerodrome Report*

**NASA** *National Aeronautics and Space Administration*

**NCAR** *National Center for Atmospheric Research*

**NCEP** *National Centers for Environmental Prediction*

**NetCDF** *Network Common Data Form*

**NLP** *Natural Language Processing*

**NOAA** *National Oceanic and Atmospheric Administration*

**NOTAM** *Notice to Airmen*

**OCR** *Optical Character Recognition*

**OKF** *Open Knowledge Foundation*

**OSN** OpenSky Network



**OWL** *Web Ontology Language*

**PANS** *Procedures for Air Navigation Services*

**QNH** *Query: Nautical Height*

**RAG** *Retrieval-Augmented Generation*

**RBAC** *Regulamentos Brasileiros da Aviação Civil*

**RBHA** *Regulamentos Brasileiros de Homologação Aeronáutica*

**RDF** *Resource Description Framework*

**REDEMET** *Rede de Meteorologia do Comando da Aeronáutica*

**SAC** *Secretaria de Aviação Civil*

**SARP** *Standards and Recommended Practices*

**SIROS** *Sistema de Registro de Operações*

**SISCEAB** *Sistema de Controle do Espaço Aéreo Brasileiro*

**SSIM** *Standard Schedules Information Manual*

**STSC** *Sistema de Tempo Severo Convectivo*

**TAF** *Terminal Aerodrome Forecast*

**TATIC** *Total Air Traffic Information Control*

**TIC** *Tecnologia da Informação e Comunicação*

**UTC** *Coordinated Universal Time*

**XML** *Extensible Markup Language*



## Lista de Figuras

|     |   |     |
|-----|---|-----|
| 2.1 | Primeiras colunas na interface do SIROS . . . . .   | 16  |
| 2.2 | Demais colunas na interface do SIROS . . . . .  | 16  |
| 2.3 | Interface de apresentação do TATIC . . . . .  | 31  |
| 2.4 | Interface de apresentação do FR24 . . . . .   | 36  |
| 3.1 | Interface de saída do INMET . . . . .   | 63  |
| 3.2 | Estações meteorológicas do INMET . . . . .  | 64  |
| 3.3 | Bucket ERA5 . . . . .   | 66  |
| 3.4 | Interface de saída NCEP-NCAR Reanalysis I . . . . .   | 69  |
| 7.1 | Arquitetura Conceitual do Sistema <b>AirData</b> . . . . .                                      | 87  |
| 7.2 | Limitações do Fluxo Tradicional de Análise de Dados . . . . .                                   | 88  |
| 7.3 | Componentes da Base de Conhecimento Semântica . . . . .   | 90  |
| 7.4 | Ontologia representando um voo da NASA . . . . .  | 92  |
| 7.5 | Exemplo de estrutura de um Grafo de Conhecimento aplicado à aviação                             | 96  |
| 7.6 | Fluxo ilustrativo da arquitetura RAG no contexto <b>AirData</b> . Fonte: Matos (2024) . . . . . | 102 |



## Lista de Tabelas

|      |   |    |
|------|---|----|
| 2.1  | Dicionário de campos da base SIROS . . . . .  | 18 |
| 2.2  | Dicionário de campos da base Microdados Básica . . . . .                            | 20 |
| 2.3  | Dicionário de campos da base Microdados Combinada . . . . .                         | 24 |
| 2.4  | Dicionário de campos da base BIMTRA . . . . .                                       | 29 |
| 2.5  | Dicionário de campos da base TATIC . . . . .  | 33 |
| 2.6  | Dicionário de campos da base Síntese RADAR . . . . .                                | 35 |
| 2.7  | Dicionário de campos da base FR24 . . . . .   | 38 |
| 2.8  | Dicionário de campos da base OSN . . . . .  | 40 |
| 2.9  | Dicionário de campos da tabela <code>state_vectors_data4</code> via Trino . . . . . | 42 |
| 2.10 | Colunas unificadas entre as bases . . . . .   | 44 |
| 2.11 | Mapeamento entre campos da base BIMTRA e TATIC . . . . .                            | 47 |
| 2.12 | Lacunas identificadas . . . . .   | 47 |
| 2.13 | Diferenças entre campos das bases BIMTRA e TATIC . . . . .                          | 48 |
| 2.14 | Formatos e granularidade das bases de dados . . . . .                               | 49 |
| 3.1  | Decodificação do METAR SBGR . . . . .   | 52 |
| 3.2  | Decodificação de uma mensagem TAF . . . . .   | 53 |
| 3.3  | Decodificação de uma mensagem SIGMET . . . . .                                      | 56 |
| 3.4  | Campos para uma requisição à API . . . . .  | 57 |
| 3.5  | Exemplo de decodificação da mensagem GAMET SBAZ. . . . .                            | 58 |
| 3.6  | Comparação entre o MAXDISPLAY e o CAPPI . . . . .                                   | 60 |



|      |   |    |
|------|---|----|
| 3.7  | Comparação entre os tipos de radares REDEMET . . . . .  | 60 |
| 3.8  | Radares Meteorológicos Disponíveis REDEMET. . . . .   | 62 |
| 3.9  | Variáveis meteorológicas do INMET . . . . .   | 63 |
| 3.10 | Resumo técnico do ERA5 . . . . .  | 65 |
| 3.11 | Comparação entre métodos de acesso ao ERA5 . . . . .  | 66 |
| 3.12 | Comparação da aplicabilidade das bases meteorológicas por fase do voo (versão transposta) . . . . . | 72 |
| 4.1  | Modelos específicos derivados do AIRM e suas aplicações . . . . .                                   | 77 |
| 4.2  | Classificação das ocorrências aeronáuticas registradas pelo CENIPA .                                | 78 |
| 4.3  | Exemplos de campos registrados pelo CENIPA para cada ocorrência .                                   | 78 |
| 4.4  | Comparação entre AISWEB e CENIPA . . . . .  | 79 |
| 5.1  | Atributos previstos para representação de normas no repositório do <b>AirData</b> . . . . .         | 81 |
| 5.2  | Comparação dos principais conjuntos normativos no contexto do <b>AirData</b>                        | 83 |
| 6.1  | As três leis de dados abertos . . . . .   | 85 |
| 6.2  | Oito princípios de dados abertos . . . . .  | 85 |
| 7.1  | Níveis de Estruturação Semântica no Domínio Aeronáutico . . . . .                                   | 93 |
| 7.2  | Etapas Iniciais da Construção da Base de Conhecimento Semântica . .                                 | 99 |



## 1 Contextualização

Este relatório apresenta o primeiro Produto do Projeto **AirData**, cujo objetivo é realizar um levantamento das bases de dados relacionadas ao transporte aéreo brasileiro e propor mecanismos para sua integração em uma Base de Conhecimento especializada. Essa base deverá ser capaz de alimentar aplicações de inteligência artificial, permitindo consultas automatizadas e análises contextualizadas para apoiar a tomada de decisão baseada em evidências.

O projeto alinha-se a objetivos estratégicos de instituições como a Agência Nacional de Aviação Civil (ANAC), o Conselho Nacional de Justiça (CNJ), o Departamento de Controle do Espaço Aéreo (DECEA) e a Secretaria de Aviação Civil (SAC), ao buscar aprimorar a governança e a interoperabilidade de informações essenciais para a segurança, eficiência e previsibilidade da malha aérea nacional. A proposta visa também favorecer o uso de tecnologias emergentes, como ontologias e grafos de conhecimento, para integrar dados de natureza operacional, meteorológica e normativa, criando uma infraestrutura de informação robusta e reutilizável.

Historicamente, as informações sobre operações de voo encontram-se fragmentadas em múltiplos sistemas mantidos por diferentes atores. O Sistema de Registro de Operações (SIROS) e os Dados Estatísticos, sob responsabilidade da ANAC, oferecem dados fundamentais para o planejamento regulatório, enquanto o Banco de Informações do Movimento de Tráfego Aéreo (BIMTRA) e o módulo *Total Air Traffic Information Control* (TATIC) do DECEA concentram informações operacionais e controle de fluxo em tempo real. Complementam esse panorama redes privadas de vigilância, como o FlightRadar24 (FR24) e a OpenSky Network (OSN), que registram posição e velocidade de aeronaves com granularidade de segundos.

A caracterização das condições meteorológicas, elemento crítico para segurança e eficiência, também depende de múltiplas fontes. Boletins *Meteorological Aerodrome Report* (METAR) e *Terminal Aerodrome Forecast* (TAF) são produzidos por estações em aeroportos estratégicos, enquanto reanálises globais como *ECMWF Reanalysis v5* (ERA5) e *National Centers for Environmental Prediction* (NCEP) oferecem cobertura uniforme. Dados do Instituto Nacional de Meteorologia (INMET) complementam esse conjunto com observações de superfície. A integração dessas informações com dados de operações viabiliza análises sobre o impacto de condições climáticas em atrasos, consumo e demanda.



A agregação dessas fontes, no entanto, enfrenta obstáculos relevantes:

- **Disparidade de terminologias:** um mesmo conceito, como número de voo, pode ser representado como “numvoo”, “callsign” ou código de empresa concatenado com número sequencial.
- **Heterogeneidade de formatos:** coexistem arquivos *Comma-Separated Values* (CSV), *Extensible Markup Language* (XML), *JavaScript Object Notation* (JSON), bancos relacionais e formatos proprietários.
- **Diferença de granularidades:** dados que variam de registros a cada quatro segundos até resumos mensais.

Essas barreiras afetam tanto a integração de dados operacionais quanto a incorporação de informações normativas — como regulamentos da ANAC, instruções do DECEA e documentos da ICAO — essenciais para contextualizar eventos e automatizar verificações de conformidade.

Para superar tais desafios, o **AirData** propõe a construção de uma base de conhecimento integrada e semanticamente enriquecida, estruturada a partir de uma ontologia aeronáutica unificada. Essa ontologia mapeará campos e conceitos de cada fonte para um modelo conceitual comum, preservando metadados críticos como ciclo *Aeronautical Information Regulation And Control* (AIRAC), procedência e histórico de versões. O uso combinado de padrões como *Resource Description Framework* (RDF) e *Web Ontology Language* (OWL) permitirá a representação formal do domínio e a construção de grafos de conhecimento que suportem raciocínio automatizado e consultas complexas.

O principal objetivo desta primeira versão do relatório é apresentar a caracterização das bases de dados avaliadas, identificar os desafios de integração e propor diretrizes iniciais para pipelines de *Extract, Transform, Load* (ETL) que alimentarão a ontologia e a base semântica do **AirData**.

A metodologia adotada estruturou-se em quatro etapas:

1. Levantamento das fontes, com análise de acessibilidade, cobertura temporal, granularidade, formato e interfaces de consulta.
2. Normalização dos dicionários de campos e identificação de divergências semânticas.



3. Definição de critérios de qualidade (completude, atualidade, coerência e interoperabilidade) alinhados ao modelo de maturidade de dados abertos de Berners-Lee (BERNERS-LEE, 2009).
4. Apontar possíveis ontologias e definição de fluxos de ETL para unificação e versionamento automático.

Nos capítulos seguintes, descrevem-se as características de cada grupo de dados: o Capítulo 2 aborda as bases de operações de voo, o Capítulo 3 trata das fontes meteorológicas, o Capítulo 4 analisa as informações aeronáuticas e o Capítulo 5 examina as bases normativas e regulamentares. O Capítulo 6 avalia a maturidade e licenciamento dos dados abertos, enquanto o Capítulo 7 apresenta a proposta conceitual e arquitetural de integração.

Ao final, espera-se que o **AirData** proporcione ganhos concretos para a gestão do transporte aéreo brasileiro, incluindo:

- Integração semântica e interoperabilidade entre sistemas.
- Redução de esforços manuais de consolidação e limpeza de dados.
- Melhoria da qualidade, confiabilidade e explicabilidade das análises.
- Suporte à tomada de decisão em tempo real com base em dados auditáveis.



## 2 Bases de Operações de Voo

Nesse capítulo serão detalhadas as principais bases de dados avaliadas como Operações de Voo no Brasil, fundamentais para a análise de atrasos e performance de voos brasileiros. São elas:

- SIROS; |
- Dados Estatísticos ANAC;
- BIMTRA;
- TATIC;
- Síntese RADAR;
- FR24;
- OSN;

As seções seguintes detalham essas bases.

### 2.1 Sistema de Registro de Operações (SIROS)

O SIROS é um sistema eletrônico da ANAC, no qual o representante da empresa aérea ou seus prepostos podem cadastrar, alterar ou excluir o registro prévio de serviços de transporte aéreo público.

No âmbito digital e eletrônico, o SIROS é uma interface direta de inserção e consulta de registros de serviços aéreos. Seu marco legal é definido pela Resolução ANAC nº 440/2017 (ANAC, 2017) e regulamentada pela Portaria 2.177/2020 (ANAC, 2020).

Além do gerenciamento dos registros de serviços de transporte aéreo, o sistema é utilizado para validação das informações inseridas de forma amostral no BIMTRA do DECEA por operadores aeroportuários.

As Figuras 2.1 e 2.2 destacam uma captura de tela do sistema SIROS para uma pesquisa correspondente ao dia 01/11/2022



| Início     | ICAO | Operador Aéreo                           | Etapa | Voo  | Equip. | Assentos | Origem | Partida Prevista |
|------------|------|--|-------|------|--------|----------|--------|------------------|
| 25/11/2024 | TAP  | TAP - TRANSPORTES AÉREOS PORTUGUESES S/A | 1     | 0009 | A21N   | 168      | LPPT   | 25/11/2024 14:15 |
| 25/11/2024 | TAP  | TAP - TRANSPORTES AÉREOS PORTUGUESES S/A | 1     | 0010 | A21N   | 168      | SBSG   | 25/11/2024 23:25 |

Figura 2.1: Primeiras colunas na interface do SIROS

| Destino | Chegada Prevista | Tipo                                 | Codeshare                    |
|---------|------------------|--------------------------------------|------------------------------|
| SBSG    | 25/11/2024 21:55 | REGULAR DE PASSAGEIROS INTERNACIONAL | UAE/4818, AZU/7102           |
| LPPT    | 26/11/2024 06:35 | REGULAR DE PASSAGEIROS INTERNACIONAL | UAE/4819, CBJ/5458, AZU/7103 |

Figura 2.2: Demais colunas na interface do SIROS

A normatização para o registro dos dados segue o padrão do *Standard Schedules Information Manual* (SSIM) publicado pela *International Air Transport Association* (IATA).

A consulta aos voos planejados pode ser feita considerando o horizonte de 365 dias. Por outro lado, o histórico dos registros de voos pode ser consultado a partir da data de 25/03/2018. O sistema possibilita a consulta por meio da aplicação dos seguintes filtros:

- Operador aéreo;
- Origem;
- Destino;
- Número do voo;
- Período da operação;
- Natureza do transporte (doméstica ou internacional);
- Objeto do transporte (carga ou passageiros).



- Serviço (regular, não regular, sobrevoos, etc.)

Ao ser inserida uma etapa de voo no sistema, automaticamente é realizada uma verificação das informações. O código de registro do voo somente é gerado se os registros inseridos forem aceitos pelo sistema. Essa etapa tem como objetivo evitar inconsistências tais como duplicação de voos com o mesmo código e registros fora do padrão SSIM.

As etapas de voo podem ser submetidas ao SIROS até o dia anterior à data de realização do serviço pretendido, sendo observadas as regras relativas à obrigatoriedade de prévio registro para a oferta dos serviços aéreos. A submissão de etapas de voos deve ser realizada no período compreendido entre às 04:00 e 23:59 horas do horário de Brasília. A partir destes registros, o sistema é atualizado diariamente às 04:00. Os campos da base de dados do SIROS podem ser consultados no Anexo 1 da (ANAC, 2020).

A Tabela 2.1 apresenta os campos da base de dados do SIROS e suas descrições.

Tabela 2.1: Dicionário de campos da base SIROS

| <b>Campo</b>      | <b>Descrição resumida</b>                  | <b>Exemplo</b>   |
|-------------------|--|--|
| ICAO              | Código ICAO da empresa aérea               | TAP  |
| Operador_Aereo    | Nome completo da empresa operadora         | TAP - TRANSPORTES AÉREOS PORTUGUESES S/A                         |
| Etapa             | Número da etapa do voo                     | 1  |
| Voo               | Número do voo                              | 0009   |
| Equip             | Código do equipamento/aeronave             | A21N   |
| Assentos          | Número de assentos da aeronave             | 168  |
| Origem            | Código ICAO do aeroporto de origem         | LPPT   |
| Aeroporto Origem  | Descrição completa do aeroporto de origem  | LISBOA - LISBOA - PORTUGAL                                       |
| Inicio            | Data de início da operação                 | 25/11/2024   |
| Partida Prevista  | Data e hora prevista da partida            | 25/11/2024 14:15   |
| Chegada Prevista  | Data e hora prevista da chegada            | 25/11/2024 21:55   |
| Tipo              | Tipo de operação aérea                     | REGULAR DE PASSAGEIROS INTERNACIONAL                             |
| Destino           | Código ICAO do aeroporto de destino        | SBSG   |
| Aeroporto Destino | Descrição completa do aeroporto de destino | GOVERNADOR ALUIZIO ALVES - SÃO GONÇALO DO AMARANTE - RN - BRASIL |
| Codeshare         | Códigos de voos compartilhados             | UAE/4818, AZU/7102   |





Por fim, os horários apresentados no SIROS, em horário oficial de Brasília ou *Coordinated Universal Time* (UTC), são relacionados aos momentos de chegada ou partida do gate (calço a calço).

## 2.2 Dados Estatísticos - Microdados

Os Dados Estatísticos do Transporte Aéreo, divulgados pela ANAC, formam a base oficial e de envio obrigatório de informações operacionais das empresas aéreas que operam no Brasil, com exceção do táxi aéreo, abrangendo operações regulares e não regulares, domésticas e internacionais. O envio é mensal, até o dia 10 do mês subsequente, por meio do sistema DataVoo, e é disciplinado pela Resolução nº 191/2011 e pelas Portarias nº 3.506/2019 (empresas brasileiras) e nº 3.507/2019 (empresas estrangeiras), que definem escopo, leiautes, prazos, regras de unicidade e de validação (ANAC, 2011, 2019b, 2019c).

Para análise em nível de registro, a base pública é disponibilizada em duas dimensões complementares. Na dimensão etapa básica, cada perna efetivamente voada é descrita desde a decolagem até o pouso subsequente, representando o estado da aeronave entre dois aeródromos. Nessa dimensão, destacam-se variáveis diretamente ligadas à execução da etapa, como identificação da aeronave (três últimas letras da matrícula), tipo de aeronave segundo designador OACI, assentos disponíveis, capacidade de carga paga, combustível consumido e horários de partida e de chegada no gate. Na dimensão etapa combinada, o foco está no objeto transportado entre o aeródromo de embarque e o de desembarque, independentemente de escalas. Ambas as bases contêm campos para passageiros pagos e gratuitos, carga paga e gratuita, correio, bagagem, trânsito doméstico e trânsito internacional (ANAC, 2023, 2019b).

A identificação consistente do voo e a integração com o SIROS devem utilizar a variável Data Prevista de Início do Voo. Todas as datas e horários referem-se ao Horário Oficial de Brasília, e o envio, manutenção e auditoria das informações ocorrem no DataVoo, cujos documentos operacionais descrevem instruções e exemplos de preenchimento coerentes com os leiautes normativos (ANAC, 2019b, 2019c, 2019a).

As Tabelas 2.2 e 2.3 apresentam os campos e descrições das bases de Dados Estatísticos - Microdados na Etapa Básica e Etapa Combinada, respectivamente.



Tabela 2.2: Dicionário de campos da base Microdados Básica

| <b>Campo</b>             | <b>Descrição resumida</b>               | <b>Exemplo</b>             |
|--------------------------|---|----------------------------|
| id_basica                | Identificador único da instância básica | 36075169                   |
| id_empresa               | Identificador numérico da empresa aérea | 1000853                    |
| sg_empresa_icao          | Código ICAO da empresa aérea            | MPH                        |
| sg_empresa_iata          | Código IATA da empresa aérea            | MP                         |
| nm_empresa               | Nome completo da empresa aérea          | MARTINAIR HOLLAND N.V.     |
| nm_pais                  | País de origem da empresa aérea         | PAÍSES BAIXOS              |
| ds_tipo_empresa          | Tipo de empresa                         | ESTRANGEIRA REGULAR        |
| nr_voo                   | Número do voo                           | 6141                       |
| nr_singular              | Sufixo do voo                           | MP                         |
| id_di                    | Identificador do tipo de voo            | 2                          |
| cd_di                    | Código do tipo de voo                   | 0                          |
| ds_di                    | Descrição do tipo de voo                | REGULAR PREVISTO EM HOTRAN |
| ds_grupo_di              | Agrupamento geral do tipo de voo        | REGULAR                    |
| dt_referencia            | Data de referência do voo               | 2025-07-03                 |
| nr_ano_referencia        | Ano da referência                       | 2025                       |
| nr_semestre_referencia   | Número do semestre                      | 2                          |
| nm_semestre_referencia   | Nome do semestre                        | 2º SEMESTRE                |
| nr_trimestre_referencia  | Número do trimestre                     | 3                          |
| nm_trimestre_referencia  | Nome do trimestre                       | 3º TRIMESTRE               |
| nr_mes_referencia        | Número do mês                           | 7                          |
| nm_mes_referencia        | Nome do mês                             | JULHO                      |
| nr_semana_referencia     | Número da semana                        | 27                         |
| nm_dia_semana_referencia | Dia da semana                           | QUINTA-FEIRA               |
| nr_dia_referencia        | Dia do mês                              | 3                          |
| nr_ano_mes_referencia    | Ano e mês no formato yyyyymm            | 202507                     |
| id_tipo_linha            | ID do tipo de linha                     | 0                          |
| cd_tipo_linha            | Código do tipo de linha                 | X                          |
| ds_tipo_linha            | Descrição do tipo de linha              | NÃO IDENTIFICADA           |
| ds_natureza_tipo_linha   | Natureza operacional da linha           | NÃO IDENTIFICADA           |



Tabela 2.2 – continuação

| <b>Campo</b>               | <b>Descrição resumida</b>                  | <b>Exemplo</b>   |
|----------------------------|--|------------------|
| ds_servico_tipo_linha      | Tipo de serviço operado                    | NÃO IDENTIFICADO |
| ds_natureza_etapa          | Classificação da etapa                     | INTERNACIONAL    |
| hr_partida_real            | Hora real de partida (Horário de Brasília) | 22:00:00         |
| dt_partida_real            | Data real da partida                       | 2025-07-03       |
| nr_ano_partida_real        | Ano da partida                             | 2025             |
| nr_semestre_partida_real   | Semestre da partida                        | 2                |
| nm_semestre_partida_real   | Nome do semestre                           | 2º SEMESTRE      |
| nr_trimestre_partida_real  | Trimestre da partida                       | 3                |
| nm_trimestre_partida_real  | Nome do trimestre                          | 3º TRIMESTRE     |
| nr_mes_partida_real        | Mês da partida                             | 7                |
| nm_mes_partida_real        | Nome do mês                                | JULHO            |
| nr_semana_partida_real     | Semana do ano                              | 27               |
| nm_dia_semana_partida_real | Dia da semana                              | QUINTA-FEIRA     |
| nr_dia_partida_real        | Dia do mês da partida                      | 3                |
| nr_ano_mes_partida_real    | Ano e mês da partida                       | 202507           |
| id_aerodromo_origem        | ID do aeródromo de origem                  | 162              |
| sg_icao_origem             | Código ICAO do aeroporto de origem         | SBKP             |
| sg_iata_origem             | Código IATA do aeroporto de origem         | VCP              |
| nm_aerodromo_origem        | Nome do aeroporto de origem                | VIRACOPOS        |
| nm_municipio_origem        | Município do aeroporto de origem           | CAMPINAS         |
| sg_uf_origem               | Unidade federativa (UF)                    | SP               |
| nm_regiao_origem           | Região brasileira                          | SUDESTE          |
| nm_pais_origem             | País de origem                             | BRASIL           |
| nm_continente_origem       | Continente do aeroporto de origem          | AMÉRICA DO SUL   |
| nr_etapa                   | Número da etapa do voo                     | 2                |
| hr_chegada_real            | Hora local da chegada real                 | 03:40:00         |
| dt_chegada_real            | Data da chegada real                       | 2025-07-04       |
| nr_ano_chegada_real        | Ano da chegada                             | 2025             |
| nr_semestre_chegada_real   | Semestre da chegada                        | 2                |
| nm_semestre_chegada_real   | Nome do semestre                           | 2º SEMESTRE      |



Tabela 2.2 – continuação

| <b>Campo</b>               | <b>Descrição resumida</b>           | <b>Exemplo</b>                  |
|----------------------------|-------------------------------------|---------------------------------|
| nr_trimestre_chegada_real  | Trimestre da chegada                | 3                               |
| nm_trimestre_chegada_real  | Nome do trimestre                   | 3º TRIMESTRE                    |
| nr_mes_chegada_real        | Mês da chegada                      | 7                               |
| nm_mes_chegada_real        | Nome do mês                         | JULHO                           |
| nr_semana_chegada_real     | Semana do ano                       | 27                              |
| nm_dia_semana_chegada_real | Dia da semana                       | SEXTA-FEIRA                     |
| nr_dia_chegada_real        | Dia do mês                          | 4                               |
| nr_ano_mes_chegada_real    | Ano e mês da chegada                | 202507                          |
| id Equipamento             | Identificador do equipamento        | 106                             |
| sg Equipamento_icao        | Código ICAO da aeronave             | B744                            |
| ds Modelo                  | Modelo da aeronave                  | BOEING 747-400                  |
| ds Matricula               | Matrícula da aeronave               | [vazio]                         |
| id Aerodromo_destino       | ID do aeródromo de destino          | 111                             |
| sg_icao_destino            | Código ICAO do aeroporto de destino | SKBO                            |
| sg_iata_destino            | Código IATA do aeroporto de destino | BOG                             |
| nm_aerodromo_destino       | Nome do aeroporto de destino        | EL DORADO INTERNATIONAL AIRPORT |
| nm_municipio_destino       | Cidade do aeroporto de destino      | BOGOTÁ                          |
| sg_uf_destino              | UF destino                          | MG                              |
| nm_regiao_destino          | Região destino                      | Sul                             |
| nm_pais_destino            | País de destino                     | COLÔMBIA                        |
| nm_continente_destino      | Continente do destino               | AMÉRICA DO SUL                  |
| nr_escala_destino          | Número da escala                    | [vazio]                         |
| lt_combustivel             | Combustível em litros               | 0                               |
| nr_assentos_ofertados      | Número de assentos oferecidos       | 0                               |
| kg_payload                 | Peso da carga útil (payload) em kg  | 110000                          |
| km_distancia               | Distância do voo em quilômetros     | 4254                            |
| nr_passag_pagos            | Passageiros pagos                   | 0                               |
| nr_passag_gratis           | Passageiros gratuitos               | 0                               |
| kg_bagagem_livre           | Peso bagagem isenta                 | 0                               |
| kg_bagagem_excesso         | Peso bagagem excesso                | 0                               |

Tabela 2.2 – continuação

| <b>Campo</b>         | <b>Descrição resumida</b>            | <b>Exemplo</b>             |
|----------------------|--------------------------------------|----------------------------|
| kg_carga_paga        | Peso da carga paga                   | 0                          |
| kg_carga_gratis      | Peso da carga gratuita               | 0                          |
| kg_correio           | Peso da carga de correio             | 0                          |
| nr_decolagem         | Número de decolagens                 | 1                          |
| nr_horas_voadas      | Horas de voo                         | 5.667                      |
| kg_peso              | Peso em quilogramas                  | 0                          |
| nr_velocidade_media  | Velocidade média                     | 750.706                    |
| nr_pax_gratis_km     | Passageiros gratuitos por quilômetro | 0                          |
| nr_carga_paga_km     | Carga paga por quilômetro            | 0                          |
| nr_carga_gratis_km   | Carga gratuita por quilômetro        | 0                          |
| nr_correio_km        | Correio por quilômetro               | 0                          |
| nr_bagagem_paga_km   | Bagagem paga por quilômetro          | 0                          |
| nr_bagagem_gratis_km | Bagagem gratuita por quilômetro      | 0                          |
| nr_ask               | Assentos oferecidos por quilômetro   | 0                          |
| nr_rpk               | Passageiros pagos por quilômetro     | 0                          |
| nr_atk               | Tonelagem oferecida por quilômetro   | 467940                     |
| nr_rtk               | Tonelagem paga por quilômetro        | 0                          |
| id_arquivo           | Identificador do arquivo de origem   | 20885                      |
| nm_arquivo           | Nome do arquivo de origem            | [vazio]                    |
| nr_linha             | Número da linha no arquivo           | [vazio]                    |
| dt_sistema           | Data/hora de inserção no sistema     | 2025-08-13 01:00:02.911387 |
| nr_chave             | Número da chave                      | 1000853614100016220250704  |





Tabela 2.3: Dicionário de campos da base Microdados Combinada

| <b>Campo</b>             | <b>Descrição resumida</b>                  | <b>Exemplo</b>             |
|--------------------------|--|----------------------------|
| id_combinada             | Identificador único da instância combinada | 95714315                   |
| id_empresa               | Identificador numérico da empresa aérea    | 1001226                    |
| sg_empresa_icao          | Código ICAO da empresa aérea               | BAW                        |
| sg_empresa_iata          | Código IATA da empresa aérea               | BA                         |
| nm_empresa               | Nome completo da empresa aérea             | BRITISH AIRWAYS PLC        |
| nm_pais                  | País de origem da empresa aérea            | INGLATERRA                 |
| ds_tipo_empresa          | Tipo de empresa                            | ESTRANGEIRA REGULAR        |
| nr_voo                   | Número do voo                              | 0246                       |
| nr_singular              | Sufixo do voo                              | AZU                        |
| id_di                    | Identificador do tipo de voo               | 2                          |
| cd_di                    | Código do tipo de voo                      | 0                          |
| ds_di                    | Descrição do tipo de voo                   | REGULAR PREVISTO EM HOTRAN |
| ds_grupo_di              | Agrupamento geral do tipo de voo           | REGULAR                    |
| dt_referencia            | Data de referência do voo                  | 2025-05-03                 |
| nr_ano_referencia        | Ano da referência                          | 2025                       |
| nr_semestre_referencia   | Número do semestre                         | 1                          |
| nm_semestre_referencia   | Nome do semestre                           | 1º SEMESTRE                |
| nr_trimestre_referencia  | Número do trimestre                        | 2                          |
| nm_trimestre_referencia  | Nome do trimestre                          | 2º TRIMESTRE               |
| nr_mes_referencia        | Número do mês                              | 5                          |
| nm_mes_referencia        | Nome do mês                                | MAIO                       |
| nr_semana_referencia     | Número da semana                           | 18                         |
| nm_dia_semana_referencia | Dia da semana                              | SÁBADO                     |
| nr_dia_referencia        | Dia do mês                                 | 3                          |
| nr_ano_mes_referencia    | Ano e mês no formato yyyyymm               | 202505                     |
| id_tipo_linha            | ID do tipo de linha                        | 0                          |
| cd_tipo_linha            | Código do tipo de linha                    | X                          |
| ds_tipo_linha            | Descrição do tipo de linha                 | NÃO IDENTIFICADA           |
| ds_natureza_tipo_linha   | Natureza operacional da linha              | NÃO IDENTIFICADA           |



Tabela 2.3 – continuação

| <b>Campo</b>               | <b>Descrição resumida</b>                  | <b>Exemplo</b>                                 |
|----------------------------|--|--|
| ds_servico_tipo_linha      | Tipo de serviço operado                    | NÃO IDENTIFICADO                               |
| ds_natureza_etapa          | Classificação da etapa                     | INTERNACIONAL                                  |
| hr_partida_real            | Hora real de partida (local)               | 15:16:00                                       |
| dt_partida_real            | Data real da partida                       | 2025-05-03                                     |
| nr_ano_partida_real        | Ano da partida                             | 2025   |
| nr_semestre_partida_real   | Semestre da partida                        | 1  |
| nm_semestre_partida_real   | Nome do semestre                           | 1º SEMESTRE                                    |
| nr_trimestre_partida_real  | Trimestre da partida                       | 2  |
| nm_trimestre_partida_real  | Nome do trimestre                          | 2º TRIMESTRE                                   |
| nr_mes_partida_real        | Mês da partida                             | 5  |
| nm_mes_partida_real        | Nome do mês                                | MAIO   |
| nr_semana_partida_real     | Semana do ano                              | 18   |
| nm_dia_semana_partida_real | Dia da semana                              | SÁBADO   |
| nr_dia_partida_real        | Dia do mês da partida                      | 3  |
| nr_ano_mes_partida_real    | Ano e mês da partida                       | 202505   |
| id_aerodromo_origem        | ID do aeródromo de origem                  | 301  |
| sg_icao_origem             | Código ICAO do aeroporto de origem         | SBGR   |
| sg_iata_origem             | Código IATA do aeroporto de origem         | GRU  |
| nm_aerodromo_origem        | Nome do aeroporto de origem                | GUARULHOS - GOVERNADOR ANDRÉ FRANCO<br>MONTORO |
| nm_municipio_origem        | Município do aeroporto de origem           | GUARULHOS                                      |
| sg_uf_origem               | Unidade federativa (UF)                    | SP   |
| nm_regiao_origem           | Região brasileira                          | SUDESTE  |
| nm_pais_origem             | País de origem                             | BRASIL   |
| nm_continente_origem       | Continente do aeroporto de origem          | AMÉRICA DO SUL                                 |
| nr_etapa                   | Número da etapa do voo                     | 1  |
| hr_chegada_real            | Hora da chegada real (Horário de Brasília) | 02:29:00                                       |
| dt_chegada_real            | Data da chegada real                       | 2025-05-04                                     |
| nr_ano_chegada_real        | Ano da chegada                             | 2025   |
| nr_semestre_chegada_real   | Semestre da chegada                        | 1  |



Tabela 2.3 – continuação

| <b>Campo</b>               | <b>Descrição resumida</b>           | <b>Exemplo</b>          |
|----------------------------|-------------------------------------|-------------------------|
| nm_semestre_chegada_real   | Nome do semestre                    | 1° SEMESTRE             |
| nr_trimestre_chegada_real  | Trimestre da chegada                | 2                       |
| nm_trimestre_chegada_real  | Nome do trimestre                   | 2° TRIMESTRE            |
| nr_mes_chegada_real        | Mês da chegada                      | 5                       |
| nm_mes_chegada_real        | Nome do mês                         | MAIO                    |
| nr_semana_chegada_real     | Semana do ano                       | 19                      |
| nm_dia_semana_chegada_real | Dia da semana                       | DOMINGO                 |
| nr_dia_chegada_real        | Dia do mês                          | 4                       |
| nr_ano_mes_chegada_real    | Ano e mês da chegada                | 202505                  |
| id_aerodromo_destino       | ID do aeródromo de destino          | 414                     |
| sg_icao_destino            | Código ICAO do aeroporto de destino | EGLL                    |
| sg_iata_destino            | Código IATA do aeroporto de destino | LHR                     |
| nm_aerodromo_destino       | Nome do aeroporto de destino        | LONDON HEATHROW AIRPORT |
| nm_municipio_destino       | Cidade do aeroporto de destino      | LONDON, ENGLAND         |
| sg_uf_destino              | UF destino                          | MG                      |
| nm_regiao_destino          | Região destino                      | SUDESTE                 |
| nm_pais_destino            | País de destino                     | REINO UNIDO             |
| nm_continente_destino      | Continente do destino               | EUROPA                  |
| nr_escala_destino          | Número da escala                    | 2                       |
| cd_cotran                  | Código COTRAN                       | 1234                    |
| ds_cotran                  | Descrição COTRAN                    | NÃO IDENTIFICADA        |
| nr_passag_pagos            | Passageiros pagos                   | 320                     |
| nr_passag_gratis           | Passageiros gratuitos               | 11                      |
| kg_bagagem_livre           | Peso bagagem isenta                 | 0                       |
| kg_bagagem_excesso         | Peso bagagem excesso                | 0                       |
| kg_carga_paga              | Peso da carga paga                  | 27398                   |
| kg_carga_gratis            | Peso da carga gratuita              | 0                       |
| kg_correio                 | Peso da carga de correio            | 0                       |
| id_arquivo                 | Identificador do arquivo de origem  | 20675                   |
| nm_arquivo                 | Nome do arquivo de origem           | NÃO IDENTIFICADO        |



Tabela 2.3 – continuação

| Campo      | Descrição resumida               | Exemplo                  |
|------------|----------------------------------|--------------------------|
| nr_linha   | Número da linha no arquivo       | NÃO IDENTIFICADO         |
| dt_sistema | Data/hora de inserção no sistema | 2025-06-15 01:00:37.1501 |



## 2.3 Banco de Informações do Movimento de Tráfego Aéreo (BIMTRA)

A base BIMTRA consolida dados operacionais detalhados de mais de 180 aeródromos brasileiros. Com alta frequência de atualização (aproximadamente 95% dos registros disponíveis em poucas horas e o restante em até um dia), essa base fornece uma visão quase em tempo real das operações aéreas no Brasil.

Os dados disponíveis no BIMTRA incluem informações como tipo de operação (decolagem ou aterrissagem), identificação do voo, matrícula da aeronave, posições de estacionamento, horários reais e previstos de movimentações, além de parâmetros de voo como altitude, rota planejada e equipamentos embarcados. A base contempla ainda variáveis específicas relacionadas à navegação aérea, como procedimentos de saída e descida, fixos de entrada/saída em *Flight Information Region* (FIR) e informações sobre órgãos *Air Traffic Services* (ATS) responsáveis pelo controle.

Além dos atributos operacionais, o BIMTRA inclui dados administrativos e técnicos como o número de pessoas a bordo, aeroportos alternativos planejados, tempo de autonomia da aeronave, registros de operadores e observações relevantes inseridas por controladores. A estrutura da base é pensada para permitir rastreabilidade total de cada movimento, sendo usada como suporte à tomada de decisão em tempo real. A Tabela 2.4 apresenta os campos da base de dados do BIMTRA e suas descrições.



Tabela 2.4: Dicionário de campos da base BIMTRA

| <b>Campo</b>  | <b>Descrição resumida</b>                             | <b>Exemplo</b>                |
|---------------|---|-------------------------------|
| deparr        | Tipo de operação realizada                            | D                             |
| numvoo        | Identificação da aeronave registrada no plano de voo  | GLO1750                       |
| matricula     | Matrícula da aeronave                                 | PPXJL                         |
| adpartida     | Aeroporto de partida                                  | SBSP                          |
| addestino     | Aeroporto de chegada                                  | SBRJ                          |
| transponder   | Código transponder SSR                                | 5442                          |
| nivelvoo      | Nível de voo ou regras como VFR/A045                  | A045                          |
| tipoaeronave  | Tipo de equipamento/aeronave                          | B738                          |
| esteiraturb   | Esteira de turbulência                                | L                             |
| tipovoo       | Tipo de voo   | S                             |
| rota          | Rota em voo conforme plano                            | REA 1909S04438W/N0170F075 DCT |
| saida         | Procedimento de saída                                 | TIL FL 110                    |
| proceddescida | Procedimento de descida                               | VFR                           |
| box           | Portão de saída/chegada ou setor operacional          | 05                            |
| fixoentrada   | Fixo de entrada                                       | BIVOV                         |
| dhfixoentrada | Data/hora da entrada no fixo                          | 2025-03-12 08:30:15:123       |
| fixosaida     | Fixo de saída (último ponto antes de sair da FIR/CTR) | ANBIX                         |
| dhfixosaida   | Data/hora de saída do fixo                            | 2025-03-12 08:30:15:123       |
| regravoo      | Regra de voo  | VI                            |
| dhinseridobd  | Data/hora de inserção no BD                           | 2025-03-12 08:30:15:123       |
| orgaoats      | Órgão operacional                                     | SBMK                          |
| regraad       | Regra de voo do aeroporto                             | I                             |
| editado       | Edição do movimento no banco de dados                 | 0 / 1                         |
| codmovimento  | Código do movimento validado                          | 57482750                      |
| codremessa    | Número do lote de remessa                             | 99832015                      |
| dhvalidado    | Data/hora do movimento validado                       | 2025-03-12 08:30:15:123       |

Continua na próxima página

Tabela 2.4 – continuação

| <b>Campo</b>   | <b>Descrição resumida</b>                   | <b>Exemplo</b>                    |
|----------------|---|-----------------------------------|
| operacao       | Código da operação                          | A                                 |
| dhmovreal      | Data/hora do movimento real (ATOT ou ALDT)  | 2025-03-12 08:30:15:123           |
| dhmovprev      | EOBT ou EIBT                                | 2025-03-12 08:30:15:123           |
| dhmovestm      | Última estimativa de movimento              | 2025-03-12 08:30:15:123           |
| tipocruzamento | Tipo de cruzamento                          | LOW                               |
| pob            | Pessoas a bordo                             | 220                               |
| altn           | Aeroporto alternativo 1                     | SBGL                              |
| altn2          | Aeroporto alternativo 2                     | SBGL                              |
| autonomia      | Tempo de autonomia da aeronave (min)        | 0230 / 0300 / 0400                |
| rmk            | Observações gerais (remarks)                | ARC08 DEST/CLARO PORCOES 500FT    |
| operador       | Nome do operador/controlador                | NRMA                              |
| rmkapp         | Observações adicionais                      | APP COORD /01 (/16JP369 /19 JP366 |
| equipamento1   | Equipamentos embarcados (TCAS, ADS-B, etc.) | SWDGIRY                           |
| equipamento2   | Equipamentos adicionais ou códigos          | C                                 |
| eet            | Tempo estimado de voo                       | DEP/NOVO ARIPUANA                 |
| outrosdados    | Dados complementares diversos               | TEXTO                             |
| numaeronaves   | Número de aeronaves                         | 2                                 |
| dac            | Número do DAC (registro aeronáutico)        | 1                                 |
| tqaligado      | Taxiway de alinhamento                      | ALPHA                             |
| taxiway        | Taxiway utilizada                           | 26 / 28,661 / 51,113              |
| numremessa     | Número de remessa                           | N0460                             |
| velocidade     | Velocidade da aeronave                      | 330                               |
| pista          | Pista do aeroporto                          | 35L                               |
| tgl            | Indicador de touch-and-go landing           | 2                                 |
| codlote        | Código do lote                              | 99832015                          |





## 2.4 Total Air Traffic Information Control (TATIC)

O TATIC é um dos módulos do sistema que controla e gerencia o fluxo de informações de tráfego aéreo existente em uma torre de controle pelo DECEA. Para tal, é utilizado um software que emprega o conceito de strips eletrônicos (DECEA, 2006). A empresa desenvolvedora do sistema do DECEA é Saipher ATC.

O funcionamento do TATIC depende das informações recebidas automaticamente da rede de mensagens *Air Traffic Service* (ATS). Por meio dessa integração, o módulo é continuamente atualizado com informações de atrasos, cancelamentos e eventos que impactam as operações das aeronaves nos aeroportos.

Outro aspecto importante na operação do TATIC é sua integração com o sistema de vigilância utilizado no Sistema de Controle do Espaço Aéreo Brasileiro (SISCEAB) - SAGITÁRIO. Esta comunicação permite a transferência dos dados, bem como a atualização das informações, permitindo a contínua prestação dos serviços de tráfego aéreo por distintos órgãos de controle.

Atualmente, o sistema conta com cobertura de mais de 60 aeroportos, representando mais de 85% do território nacional e mais de 95% da aviação comercial. Os dados são disponibilizados em tempo real, com cobertura dos aeródromos mais movimentados (ATC, 2020). A Figura 2.3 apresenta a visualização do TATIC, com destaque para as *Electronic Flight Progression Strips* (EFPS), que são a base das operações da ferramenta.

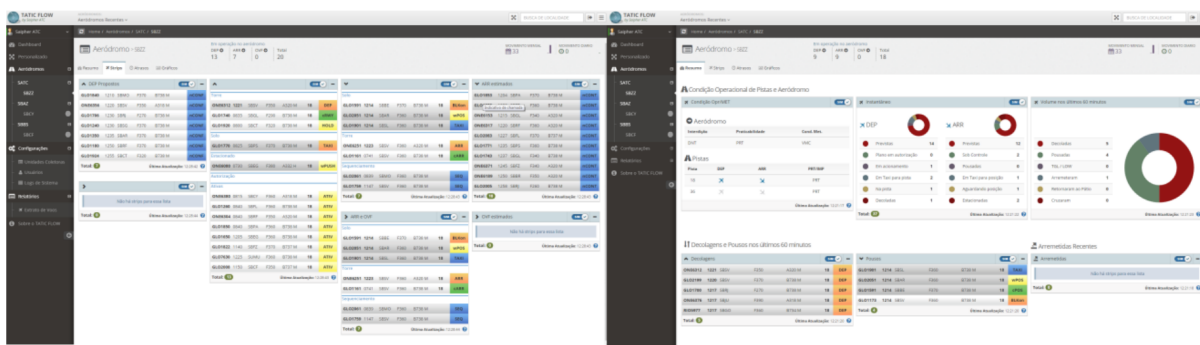


Figura 2.3: Interface de apresentação do TATIC

Como o módulo é automatizado, é criado um banco de dados que apresenta informações operacionais em tempo real que é compartilhado para o módulo TATIC de outras unidades. Ressalta-se que as informações armazenadas correspondem aos comandos inseridos pelos controladores de tráfego aéreo durante as operações, sendo destacados os principais eventos:



- Solicitação e autorização de acionamento da aeronave;
- Solicitação e autorização do pushback;
- Solicitação e autorização de taxiamento;
- Ingresso nas posições 2 e 3;
- Autorização e realização da decolagem;
- Autorização e realização do pouso;
- Estacionamento da aeronave;
- Indicação de pista livre.

Se houver tráfego que possa interferir, a aeronave que vai partir será mantida na posição 2 a 90° com a direção de pouso. Normalmente, nessa posição serão testados os motores. Quando duas ou mais aeronaves atingirem a posição 2, deverão manter-se a 45° com a direção de pouso. A autorização para decolagem será dada na posição 3, se não foi possível fazê-lo na posição 2. A fonte dessas notas é ICA 100-37 (DECEA, 2024).

Os registros dos horários das operações aeroportuárias na base TATIC são em fuso UTC, enquanto nas demais bases de dados é empregado o horário local.

Uma das vantagens da base de dados TATIC é a apresentação de informações adicionais em comparação às demais. Neste sentido, são descritos os horários de todos os procedimentos em qualquer fase do voo com informações específicas do ponto de vista de controle do tráfego aéreo. A Tabela 2.5 apresenta os campos da base de dados TATIC e suas descrições.

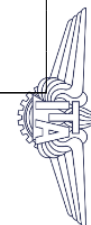


Tabela 2.5: Dicionário de campos da base TATIC

| <b>Campo</b> | <b>Descrição resumida</b>  | <b>Exemplo</b>                       |
|--------------|--|--------------------------------------|
| id           | Identificador único da instância   | B16AF9F7-617C-4E62-B6C6-A1640B6C8934 |
| flowid       | Identificador do fluxo associado   | a9179e09-dc33-4f74-9199-0d5e857b0ca2 |
| createdat    | Data e hora de criação do registro   | 2025-04-01 00:35:03.213              |
| locality     | Localidade ou aeroporto de controle local  | SBGR                                 |
| callsign     | Identificador do voo (prefixo + número)  | TAM4536                              |
| adep         | Aeroporto de origem  | SBGR                                 |
| ades         | Aeroporto de destino   | SBJP                                 |
| rv           | Regras de voo  | IV                                   |
| flighttype   | Tipo de voo  | S                                    |
| runway       | Cabeceira da pista utilizada   | 10L                                  |
| equipment    | Equipamentos embarcados (códigos ICAO)   | SWDZGHIRXY/S                         |
| acfttype     | Tipo de aeronave   | A321                                 |
| eventtype    | Tipo de evento monitorado  | DEP                                  |
| eobt         | Hora estimada de início de taxi (estimated off-block time)                         | 2025-03-31 23:50:00.000              |
| wpush        | Horário estimado de pushback (esperado)  | 2025-03-31 23:51:46.857              |
| cpush        | Horário de clearance (autorização) do pushback                                     | 2025-03-31 23:51:52.760              |
| wtaxi        | Horário estimado de início do táxiamento   | 2025-03-31 23:58:54.977              |
| taxi         | Horário confirmado de início do táxiamento   | 2025-03-31 23:58:57.957              |
| hold         | Horário de entrada em espera (hold) se aplicável                                   | 2025-04-01 00:01:36.440              |
| crwy         | Horário estimado de entrada na pista   | 2025-04-01 00:01:37.300              |
| cdep         | Horário confirmado de decolagem  | 2025-04-01 00:02:14.320              |
| dep          | Horário final de decolagem registrado (ATOT) ou tempo confirmado de saída do setor | 2025-04-01 00:03:22.240              |
| eta          | Estimated Time of Arrival (hora estimada de chegada)                               | 2025-03-31 23:44:00.000              |
| arr          | Horário confirmado de chegada (ALDT)   | 2025-03-31 23:23:58.770              |
| cpos         | Horário confirmado de ocupação de posição (gate ou posição remota)                 | 2025-03-31 23:27:53.033              |
| transponder  | Código transponder/SSR da aeronave   | 4253                                 |



## 2.5 Síntese RADAR

A base Síntese RADAR é composta por registros provenientes dos sensores de vigilância do espaço aéreo brasileiro, coletados em intervalos regulares de aproximadamente quatro segundos. É utilizado principalmente para a reconstrução precisa de trajetórias de aeronaves em voo, com potencial para análises operacionais detalhadas.

Cada entrada na base representa uma detecção individual, contendo informações como posição geográfica (latitude e longitude), velocidade, nível de voo, código do transponder, bem como metadados técnicos sobre o tipo e a fonte do detector. A estrutura permite distinguir entre trajetórias reais e estimadas, além de identificar segmentos de espera ou padrão de órbita próximos a aeroportos, o que é fundamental para análises de capacidade e gerenciamento de fluxo de tráfego.

Por ser uma base de alta granularidade, seu uso demanda grande capacidade de processamento e armazenamento. No contexto deste projeto, a base RADAR é utilizada para compor as sequências temporais de voo, que posteriormente são correlacionadas com informações operacionais e regulatórias oriundas de outras fontes (como BINTRA, TACTIC, ANAC), permitindo uma visão integrada e confiável do comportamento das aeronaves ao longo de suas trajetórias. A Tabela 2.6 mostra os campos da base de dados Síntese RADAR e suas descrições.



Tabela 2.6: Dicionário de campos da base Síntese RADAR

| <b>Campo</b>      | <b>Descrição resumida</b>                                    | <b>Exemplo</b>          |
|-------------------|--|-------------------------|
| id                | Identificador interno dos registros de RADAR                 | 10746903083             |
| nr_id             | Identificador interno dos registros de RADAR                 | 98                      |
| nr_seq            | Identificador interno dos registros de RADAR                 | 804                     |
| nr_cycle          | Ciclo de consolidação do dado (geralmente a cada 4 segundos) | 3                       |
| ds_acctypes       | Região de cobertura do RADAR responsável pela medição        | ['CURITIBA']            |
| dh_epoch          | Timestamp em milissegundos desde a época Unix (epoch)        | 1743508801264           |
| ds_detectorSource | Fonte de detecção  | ASSOCIADO_2_D           |
| ds_pathType       | Tipo de trajetória registrada                                | REAL                    |
| vl_latitude       | Posição geográfica do alvo em latitude                       | -20.130890              |
| vl_longitude      | Posição geográfica do alvo em longitude                      | -54.608133              |
| nr_speed          | Velocidade instantânea                                       | 109 kt                  |
| nr_speedBearing   | Direção do movimento (azimute)                               | 5.86°                   |
| nr_ssr            | Código transponder SSR                                       | 115                     |
| nr_flightLevel    | Nível de voo em pés  | 46                      |
| ds_registration   | Matrícula da aeronave (nulo quando não identificado)         | NULL                    |
| dt_radar          | Data e hora da geração do registro de RADAR                  | 2025-04-01 12:00:01.000 |



## 2.6 FlighRadar24 (FR24)

O FR24 é um serviço online de rastreamento de voos disponibilizado para o público geral. A Figura 2.4 apresenta a tela de abertura do FR24. O acesso à ferramenta pode ser feito em FR24 (2025).

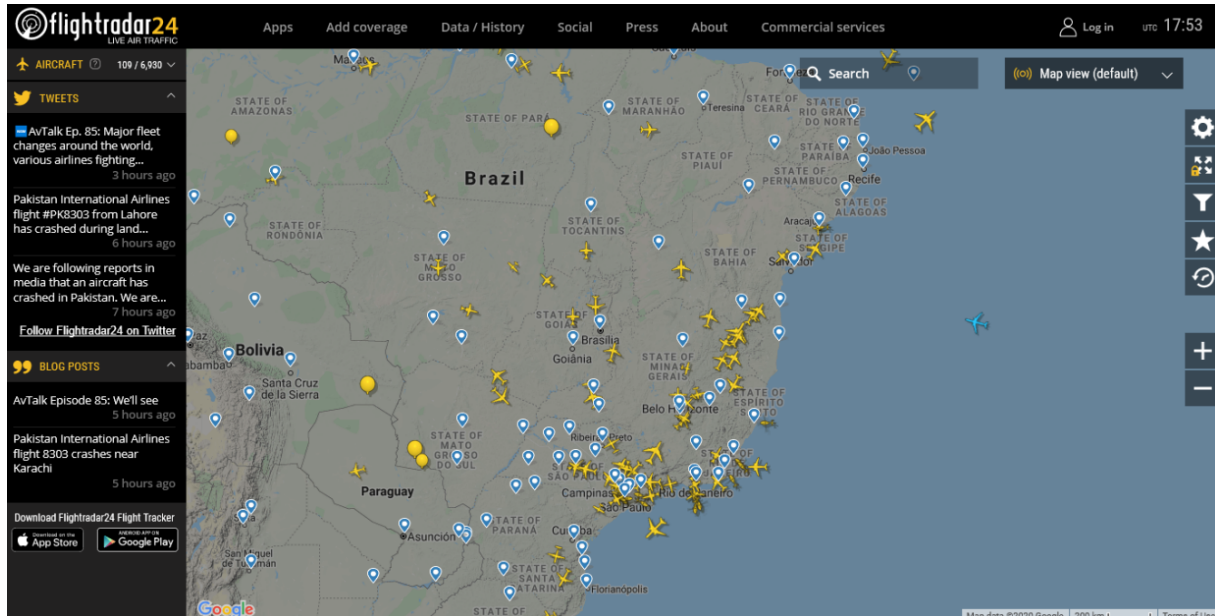


Figura 2.4: Interface de apresentação do FR24

Esse serviço possui uma grande rede de receptores *Automatic Dependent Surveillance-Broadcast* (ADS-B) pelo mundo (constituída a partir de *crowdsourcing*) que recebem registros de posicionamento e identificação de aeronaves apropriadamente equipadas e enviam essas informações para seus servidores, viabilizando, desta forma, o rastreamento de aeronaves no espaço aéreo em tempo real.

Os registros podem ser extraídos do site do FR24 por meio de *Web Scraping*, que consiste no uso de programas de computadores automatizados que conseguem extrair os registros de interesse da web, utilizando diferentes bibliotecas computacionais.

Por outro lado, esta técnica torna o sistema suscetível à interrupção de fornecimento devido a intervenções no site pelo provedor. Os dados que podem ser obtidos do FR24 contém informações de georreferenciamento (latitude, longitude e altitude) e performance (direção e velocidade) das aeronaves e também o instante de tempo no qual uma etapa estratégica do voo foi realizada, tais como: procedimento de decolagem, transição para a etapa de cruzeiro e procedimento de pouso.

É importante destacar que o uso de técnicas de web scraping para coleta de dados



do FR24 é restrito e configura violação dos termos e condições de uso do serviço. Considerando que a lista de bases apresentada neste documento não é exaustiva, a utilização da base FR24 será analisada conforme a necessidade, bem como a flexibilidade para inclusão ou exclusão de dados.

É importante frisar que o FR24 não disponibiliza os horários de saída e chegada no gate (calço a calço). Como regra geral, os horários registrados pelo FR24 correspondem aos momentos de toque na pista e de decolagem. Esses dados são registrados em horário local. A Tabela 2.7 mostra os campos da base de dados FR24 e suas descrições.



Tabela 2.7: Dicionário de campos da base FR24

| <b>Campo</b> | <b>Descrição resumida</b>  | <b>Exemplo</b> |
|--------------|--|----------------|
| flight_id    | Identificador numérico individual alocado pelo FR24 para cada voo      | 123456789      |
| aircraft_id  | Identificador numérico individual alocado pelo FR24 para cada aeronave | 987654321      |
| reg          | Código alfanumérico que detalha o registro da aeronave                 | PR-GTS         |
| equip        | Código alfanumérico que detalha o modelo da aeronave                   | B738           |
| callsign     | Indicativo ICAO do voo   | GLO1234        |
| flight       | Indicativo IATA do voo   | G31234         |
| sched_from   | Código IATA do aeroporto de partida                                    | GRU            |
| sched_to     | Código IATA do aeroporto de chegada detalhado no plano de voo          | CGH            |
| real_to      | Código IATA do aeroporto de chegada detectado pelo ADS-B               | CGH            |
| reserved     | Informação reservada ao FR24 (sempre com informação nula)              | -              |
| snapshot_id  | Horário da coleta da informação em formato unix                        | 1617187200     |
| altitude     | Altitude da aeronave em pés (ft)                                       | 35000          |
| heading      | Direção da aeronave em relação ao norte magnético (graus)              | 270            |
| latitude     | Latitude em coordenada UTM   | -23.435        |
| longitude    | Longitude em coordenada UTM  | -46.473        |
| RADAR_id     | Identificador numérico individual alocado pelo FR24 para a antena      | 1234           |
| speed        | Velocidade da aeronave em relação ao chão (nós - kts)                  | 460            |
| squawk       | Código numérico com a resposta do transponder para o órgão ATC         | 7500           |



## 2.7 OpenSky Network (OSN)

O OSN coleta dados de tráfego aéreo em tempo real a partir de sinais ADS-B e mensagens Mode S captadas por uma rede de receptores distribuídos globalmente. Cada aeronave equipada com transponder ADS-B transmite periodicamente informações sobre sua posição, velocidade, altitude, entre outras.

Esses dados são reunidos em registros chamados vetores de estado, que sintetizam o estado da aeronave em um determinado momento, incluindo sua localização, velocidade, altitude, código ICAO e *callsign*.

Os dados do OSN podem ser acessados por:

- **REST API:** permite recuperar vetores de estado, trajetórias e informações de voo ao vivo ou históricas para pesquisas e fins não comerciais.
- **Consulta SQL via Trino:** para usuários que integram o OpenSky com ferramentas analíticas, é possível consultar os dados armazenados no *data lake* do OSN usando a linguagem SQL via Trino, facilitando análises mais customizadas e agregadas.

O OSN não fornece dados comerciais como horários de voos ou atrasos, apenas dados derivados diretamente dos sinais ADS-B. Cada vetor de estado representa a aeronave em um instante, contendo dados de posição, velocidade e identificação.

A cobertura de dados do OSN no Brasil ainda é bastante limitada, em razão do reduzido número de antenas receptoras disponíveis na plataforma de código aberto. No entanto, espera-se que a integração com os dados provenientes das bases RADAR e/ou FR24 complemente e melhore a qualidade das informações de trajetória utilizadas no projeto.

Dados de posição e velocidade são considerados válidos se atualizados nos últimos 15 segundos, caso contrário podem estar ausentes ou serem omitidos. Além disso, o sistema identifica aeronaves pelo endereço ICAO único de seu transponder. A Tabela 2.8 mostra os campos da base de dados OSN via REST API e suas descrições.

Tabela 2.8: Dicionário de campos da base OSN

| <b>Campo</b>    | <b>Descrição resumida</b>   | <b>Exemplo</b>  |
|-----------------|---|-----------------|
| icao24          | Endereço ICAO 24-bit do transponder da aeronave (hexadecimal)       | c0ffee          |
| callsign        | Código de identificação do voo (até 8 caracteres), pode ser nulo    | GLO1234         |
| origin_country  | País de origem da aeronave com base no ICAO24                       | Brazil          |
| time_position   | Timestamp (Unix) da última atualização de posição conhecida         | 1617187200      |
| last_contact    | Timestamp (Unix) da última mensagem recebida                        | 1617187215      |
| longitude       | Longitude em coordenadas WGS-84 (graus decimais)                    | -46.625290      |
| latitude        | Latitude em coordenadas WGS-84 (graus decimais)                     | -23.533773      |
| baro_altitude   | Altitude barométrica (em metros), derivada da pressão do ar         | 11277.6         |
| on_ground       | Indica se a aeronave está no solo (TRUE/FALSE)                      | FALSE           |
| velocity        | Velocidade sobre o solo (em metros por segundo)                     | 231.1           |
| true_track      | Direção em relação ao norte verdadeiro (graus)                      | 87.5            |
| vertical_rate   | Taxa de subida ou descida (em m/s) – positiva = subida              | -2.5            |
| sensors         | Lista de IDs dos sensores receptores que contribuíram com os dados  | [1123, 1456]    |
| geo_altitude    | Altitude geométrica da aeronave (em metros), se disponível          | 11305.8         |
| squawk          | Código numérico de 4 dígitos do transponder (Squawk), se disponível | 7500            |
| spi             | Flag de intenção especial (Special Purpose Indicator)               | TRUE            |
| position_source | Fonte da posição: 0=ADS-B, 1=ASTERIX, 2=MLAT, 3=FLARM               | 0               |
| category        | Categoria da aeronave (0 a 15), conforme especificação ICAO         | 6 (peso pesado) |





Além da API pública, o OSN disponibiliza acesso completo aos seus dados por meio de uma interface de consulta SQL (Trino), integrada ao armazenamento MinIO. Este serviço é destinado a pesquisadores vinculados a universidades, órgãos governamentais e autoridades aeronáuticas para fins de pesquisa ou investigação de ocorrências. Para uso comercial ou privado, é necessário obter uma licença específica.

A tabela de maior interesse é a `state_vectors_data4`, que contém os vetores de estado das aeronaves rastreadas por ADS-B. Cada linha representa um snapshot do estado da aeronave em um dado momento.

A Tabela 2.9 mostra os campos da tabela de vetores de estado coletados via SQL com Trino.

Tabela 2.9: Dicionário de campos da tabela state\_vectors\_data4 via Trino

| <b>Campo</b>  | <b>Descrição resumida</b>                                  | <b>Exemplo</b> |
|---------------|--|----------------|
| time          | Timestamp UNIX referente ao momento do vetor de estado     | 1480760792     |
| icao24        | Identificador ICAO de 24 bits da aeronave                  | a0d724         |
| lat           | Latitude da posição mais recente (WGS84)                   | 37.8946        |
| lon           | Longitude da posição mais recente (WGS84)                  | -88.9333       |
| velocity      | Velocidade sobre o solo (m/s)                              | 190.85         |
| heading       | Rumo da aeronave (graus a partir do norte geográfico)      | 265.83         |
| vertrate      | Taxa de subida/descida (m/s)                               | 0              |
| callsign      | Identificador de voo transmitido pela aeronave             | UPS858         |
| onground      | Indicador se a aeronave está no solo                       | false          |
| alert         | Indicador de emergência (alert squawk)                     | false          |
| spi           | Indicador de posição especial (Special Position Indicator) | false          |
| squawk        | Código transponder de 4 dígitos atribuído pelo ATC         | 7775           |
| baroaltitude  | Altitude barométrica (m)                                   | 9144           |
| geoaltitude   | Altitude geométrica via GNSS (m)                           | 9342.12        |
| lastposupdate | Timestamp da última atualização de posição                 | 1480760704.36  |
| lastcontact   | Último momento de contato com o receptor OpenSky           | 1480760709.99  |
| serials       | Lista de identificadores de sensores ADS-B                 | [4201, 10432]  |
| hour          | Hora inicial do bloco de dados (timestamp UNIX)            | 1480759200     |





## 2.8 Desafios de Integração entre Bases de Dados

A integração entre as diferentes bases de dados aeronáuticos que serão utilizadas no projeto **AirData** representa um desafio considerável, principalmente devido à diversidade de estruturas, nomenclaturas e granularidades envolvidas. Cada fonte de dados foi originalmente concebida com propósitos distintos (e.g., monitoramento tático, planejamento regulatório, controle operacional, vigilância aérea), o que implica em variações nos campos disponíveis, na frequência de atualização e nas unidades ou formatos de representação.

Um dos principais obstáculos é a presença de campos que, apesar de apresentarem nomes diferentes, carregam significados equivalentes ou complementares. Por exemplo, o número do voo pode ser registrado como `numvoo`, `callsign` ou mesmo de forma combinada a partir de uma coluna com códigos de empresa e outra com os números dos voos. Problemas similares ocorrem com informações como horários de partida e chegada, modelos de aeronave, aeródromos de origem/destino e regras de voo.

Diante dessas disparidades, foi iniciado um processo sistemático de mapeamento semântico entre os campos das diferentes fontes. Esse trabalho envolve a definição de um conjunto de *campos unificados*, aos quais cada variável original é associada conforme seu significado. Isso servirá de base para a definição da ontologia que o modelo de LLM usará para dar significado semântico aos dados brutos. Essa abordagem permite construir um modelo lógico comum, possibilitando análises comparáveis e consolidadas.

A Tabela 2.10 ilustra os principais campos unificados definidos até o momento, com destaque para os mapeamentos entre BIMTRA, TATIC, RADAR, e Dados Estatísticos. Além dos nomes e origens, são indicadas observações sobre as transformações necessárias ou casos de ambiguidade. Este mapeamento será expandido e refinado ao longo do projeto.

Tabela 2.10: Colunas unificadas entre as bases

| <b>Campo unificado</b>     | <b>Bases e Colunas</b>   | <b>Observações / Exemplos</b>                          |
|----------------------------|--|--|
| callsign                   | BIMTRA: numvoo<br>TATIC: callsign<br>Dados Estatísticos: nr_voo + sg_empresa_icao<br>RADAR: ds_registration          | Combinação de número do voo e código ICAO ou matrícula |
| aeródromo_origem           | BIMTRA: adpartida<br>TATIC: adept<br>Dados Estatísticos: sg_icao_origem<br>RADAR: ausente                            | Código ICAO do aeroporto de origem                     |
| aeródromo_destino          | BIMTRA: addestino<br>TATIC: ades<br>Dados Estatísticos: sg_icao_destino<br>RADAR: ausente                            | Código ICAO do aeroporto de destino                    |
| partida_real               | BIMTRA: operacao<br>TATIC: dep<br>Dados Estatísticos: dt_partida_real + hr_partida_real<br>RADAR: dt_radar (parcial) | Data e hora reais de partida                           |
| chegada_real               | BIMTRA: operacao<br>TATIC: arr<br>Dados Estatísticos: dt_chegada_real + hr_chegada_real<br>RADAR: não aplicável      | Data e hora reais de chegada                           |
| codigo_di                  | BIMTRA: ausente<br>TATIC: ausente<br>Dados Estatísticos: ds_di<br>RADAR: ausente                                     | Código do tipo de operação                             |
| tipo_voo                   | BIMTRA: tipovoo<br>TATIC: flighttype<br>Dados Estatísticos: ausente<br>RADAR: ausente                                | Tipo do voo (scheduled, general, etc.)                 |
| Continua na próxima página |  |  |





Tabela 2.10 – continuação

| <b>Campo unificado</b>     | <b>Bases e Colunas</b>   | <b>Observações / Exemplos</b>  |
|----------------------------|--|--------------------------------|
| transponder                | BIMTRA: transponder<br>TATIC: transponder<br>Dados Estatísticos: ausente<br>RADAR: nr_ssr                        | Código transponder SSR         |
| modelo_aeronave            | BIMTRA: tipoaeronave<br>TATIC: acfttype<br>MICRODADOS BÁSICA: ds_modelo<br>RADAR: ausente<br>SIROS: Equip.       | Modelo da aeronave             |
| regra_voo                  | BIMTRA: regravoo<br>TATIC: rv<br>Dados Estatísticos: ausente<br>RADAR: ausente                                   | Regras de voo (VFR, IFR, etc.) |
| pista                      | BIMTRA: pista<br>TATIC: runway<br>Dados Estatísticos: ausente<br>VRA: ausente<br>RADAR: ausente                  | Cabeceira da pista             |
| equipamento                | BIMTRA: equipamento1<br>TATIC: equipment<br>Dados Estatísticos: ausente<br>RADAR: ausente                        | Equipamentos embarcados        |
| actual_departure_time      | BIMTRA: dhmovreal<br>TATIC: dep, cdep<br>Dados Estatísticos: dt_partida_real + hr_partida_real<br>RADAR: ausente | Hora real de partida           |
| actual_landing_time        | BIMTRA: dhmovreal<br>TATIC: arr<br>Dados Estatísticos: dt_chegada_real + hr_chegada_real<br>RADAR: ausente       | Hora real de chegada           |
| Continua na próxima página |  |                                |



Tabela 2.10 – continuação

| <b>Campo unificado</b>   | <b>Bases e Colunas</b>  | <b>Observações / Exemplos</b>       |
|--------------------------|---|-------------------------------------|
| estimated_departure_time | BIMTRA: dhmovprev<br>TATIC: eobt<br>Dados Estatísticos: ausente<br>RADAR: ausente                         | Hora estimada de partida            |
| eventtype                | BIMTRA: deparr<br>TATIC: eventtype<br>Dados Estatísticos: ausente<br>RADAR: ausente                       | Tipo de evento (decolagem, chegada) |
| pob                      | BIMTRA: pob<br>TATIC: ausente<br>Dados Estatísticos: nr_passag_pagos + nr_passag_gratis<br>RADAR: ausente | Pessoas a bordo                     |
| nivel_voo                | BIMTRA: nivelvoo<br>TATIC: ausente<br>Dados Estatísticos: ausente<br>RADAR: nr_flightLevel                | Nível de voo                        |



Durante o processo de integração, observam-se também situações de sobreposição semântica parcial. Certos campos aparentemente equivalentes apresentam variações conceituais sutis, como é o caso das colunas *box* e *cpos*, ou de tempos estimados de voo (*eet* e *eta*). Essas discrepâncias requerem validação com especialistas operacionais e, em alguns casos, aplicação de regras de negócio específicas para correção ou alinhamento. A Tabela 2.11 apresenta algumas dessas sobreposições vistas.

Tabela 2.11: Mapeamento entre campos da base BIMTRA e TATIC

| BIMTRA | TATIC | Observações  |
|--------|-------|--|
| box    | cpos  | Posição operacional (gate) aproximável, mas a semântica pode divergir.           |
| eet    | eta   | EET (tempo estimado de voo) pode ser aproximado pela diferença entre EOBT e ETA. |

Outra categoria relevante de desafio são as lacunas entre conjuntos. Algumas informações estão disponíveis apenas em uma base específica, como é o caso de características econômicas e geográficas nos Microdados Combinados, dados em tempo real na base Síntese RADAR, ou status de voo e capacidade nominal presentes apenas na VRA. Essas assimetrias demandam decisões sobre preenchimento, descarte ou incorporação seletiva conforme a finalidade analítica. A Tabela 2.12 apresenta algumas lacunas encontradas.

Tabela 2.12: Lacunas identificadas

| Categoria                     | Itens Exclusivos  |
|-------------------------------|---|
| Exclusivos BIMTRA             | nivelvoo; fixoentrada, fixosaida; autonomia, pob, altn; rmk, rmkapp, outrosdados; dac, numaeronaves, tqaligado, taxiway, tgl  |
| Exclusivos TATIC              | wpush, cpush, wtaxi, taxi, hold, cdep, cpos; flowid, id, createdat  |
| Exclusivos Dados Estatísticos | Informações econômico-operacionais: kg_bagagem_, kg_carga_, kg_correio; Dados geográficos: UF, município, região, continente; Segmentação temporal refinada: ano, mês, semana, trimestre, semestre; Dados regulatórios: cotran, ds_di |
| Continua na próxima página    |   |



**Tabela 2.12 – continuação**

| <b>Categoria</b> | <b>Itens Exclusivos</b>   |
|------------------|---|
| Exclusivos RADAR | Posição em tempo real (lat/lon), velocidade e heading; Detecção e tipo de trajetória (ds_detectorSource, ds_pathType); Altitude real (nr_flightLevel) |

Além das questões de nomenclatura, há diferenças relevantes nas definições semânticas e nos formatos adotados por cada base. Exemplos incluem o significado de eventos como decolagem ou chegada, o tipo de voo (regular, geral, militar), ou a codificação de regras de voo.

Adicionalmente, formatos de data/hora variam entre padrões ISO, timestamps Unix Epoch (que é definido como um número de segundos decorridos desde 1º de janeiro de 1970) e combinações de colunas distintas. Essa heterogeneidade requer um processo robusto de padronização e enriquecimento semântico. A Tabela 2.13 mostra algumas dessas diferenças nas definições semânticas.

**Tabela 2.13: Diferenças entre campos das bases BIMTRA e TATIC**

| <b>Diferença</b>                       | <b>Impacto / Exemplo</b>  |
|--|---|
| saida (BIMTRA) vs crwy (TATIC)         | BIMTRA traz SID ou ponto inicial, TATIC foca no timestamp estimado de entrada na pista. |
| tipovoo (BIMTRA) vs flighttype (TATIC) | BIMTRA pode conter “M”, “X”; TATIC pode omitir ou tratar de modo diferente.             |
| deparr vs eventtype                    | Mapeamento simples mas com nomes diferentes.  |
| regraad vs regravoo (BIMTRA)           | Podem representar regras de voo distintas: uma da origem, outra da aeronave.            |

Por fim, a granularidade temporal também impõe limitações: enquanto o RADAR opera em altíssima frequência (4 segundos), outras fontes fornecem registros agregados por voo. A integração dessas escalas distintas exige estratégias robustas, sempre com controle de qualidade e rastreabilidade. A Tabela 2.14 mostra algumas diferenças de granularidade nas bases.



Tabela 2.14: Formatos e granularidade das bases de dados

| Aspecto                    | Descrição / Exemplo   |
|----------------------------|---|
| Formato de data/hora       | <b>BIMTRA, TATIC:</b> yyyy-mm-dd hh:mm:ss.mmm<br><b>RADAR:</b> 1743508801264<br><b>MICRODADOS:</b> yyyy-mm-dd, hh:mm:ss<br><b>VRA:</b> dd/mm/yyyy hh:mm |
| Codificação de tipo de voo | <b>BIMTRA:</b> S, G, N, M, X<br><b>TATIC:</b> S, G, N<br><b>MICRODADOS:</b> descrições extensas   |
| Granularidade temporal     | <b>RADAR:</b> registros a cada 4 segundos<br><b>TATIC:</b> eventos pontuais<br><b>Demais bases:</b> agregados por voo                                   |

## 2.9 Considerações Finais

Em síntese, as bases de operações de voo analisadas neste capítulo constituem o núcleo informacional sobre o qual o Projeto **AirData** poderá construir análises operacionais, diagnósticos e previsões. Embora cada sistema apresente objetivos, granularidades e formatos próprios, todos contribuem de forma complementar para a compreensão do ciclo de vida de um voo, desde o planejamento até a execução e registro final.

Portanto, a integração coerente desses conjuntos — respeitando suas particularidades semânticas, temporais e estruturais — será importante para viabilizar consultas unificadas, indicadores confiáveis e aplicações de inteligência artificial voltadas à melhoria da eficiência da aviação civil brasileira.



### 3 Bases de Dados Meteorológicos

Neste capítulo, são descritas as fontes de bases de dados meteorológicos avaliadas para a construção da base de conhecimento, a citar:

- Rede de Meteorologia do Comando da Aeronáutica (REDEMET);
- INMET;
- ERA5;
- NCEP–NCAR Reanalysis I e NCEP/DOE Reanalysis II.

Cada uma das bases tem seu próprio foco, modo de coleta de dados e organização de informações. Além disso, elas se complementam, de forma que uma união desses dados proporcionará uma visão mais ampla das situações meteorológicas de cada local.

#### 3.1 REDEMET

Nesta seção, encontram-se os produtos disponibilizados pela Rede de Meteorologia do Comando da Aeronáutica (REDEMET) para a obtenção de dados meteorológicos de diversos formatos e cada um com seu objetivo específico. Eles são descritos na seguinte ordem:

- Meteorological Aerodrome Report (METAR)
- Terminal Aerodrome Forecast (TAF)
- Sistema de Tempo Severo Convectivo (*Sistema de Tempo Severo Convectivo* (STSC))
- SIGMET
- GAMET
- RADAR



Destaca-se que para o acesso a *Application Programming Interface* (API) do REDEMET, que contém todos os dados dessas bases é necessário uma aprovação prévia do DECEA.

### 3.1.1 *Meteorological Aerodrome Report (METAR)*

O METAR é um informe codificado, associado às observações meteorológicas na superfície. É utilizado para receber informações sobre condições do tempo em um aeródromo específico, por meio de mensagens. Nelas, encontram-se informações sobre:

- Indicador do aeródromo (código *International Civil Aviation Organization* (ICAO));
- Data e horário da observação (UTC);
- Direção e intensidade do vento;
- Visibilidade horizontal;
- Fenômenos meteorológicos significativos (chuva, névoa, trovoadas);
- Cobertura e tipo de nuvens;
- Temperatura e ponto de orvalho;
- Pressão atmosférica reduzida ao nível do mar (*Query: Nautical Height* (QNH));  
e
- Informações suplementares relevantes (wind shear, runway state, etc.).

Por exemplo, uma mensagem METAR pode ser apresentada da seguinte forma:

SBGR 231400Z 12005KT 8000 SCT020 24/20 Q1018

A Tabela 3.1 apresenta a decodificação detalhada dessa mensagem, permitindo interpretar cada elemento de forma clara e objetiva.



Tabela 3.1: Decodificação do METAR SBGR

| Elemento                | Descrição   |
|-------------------------|---|
| Código ICAO             | SBGR — Aeroporto Internacional de São Paulo/Guarulhos |
| Data/Hora               | 23 do mês às 14:00 UTC                                |
| Vento                   | Direção 120° a 5 nós                                  |
| Visibilidade            | 8 km  |
| Cobertura de nuvens     | Nuvens espalhadas (SCT) a 2000 pés                    |
| Temperatura             | 24 °C   |
| Ponto de orvalho        | 20 °C   |
| Pressão ao nível do mar | 1018 hPa  |

Os boletins apresentam uma observação por hora. Assim, para o cruzamento desses dados com dados de operações de voo, é preciso sempre pegar os dados com o horário mais próximo, tanto de chegada quanto de partida do voo, o que compromete a precisão da análise.

O acesso a esses dados pode ser realizado por meio de uma API da REDEMETS, mantida pelo DECEA. Porém, para criar requisições à API, é necessária a permissão prévia do DECEA.

Outra opção é um pacote da linguagem de programação R chamado *pmetar* (CWIEK, 2025) que fornece seus dados de forma gratuita e sem a necessidade de um registro, pela Universidade Estadual de Iowa. Basta fornecer a data de início e de fim do período e o código ICAO.

### 3.1.2 Terminal Aerodrome Forecast (TAF)

O código TAF é a previsão das condições meteorológicas em um aeródromo durante o período de validade, incluindo qualquer mudança considerada significativa para as operações aéreas. Ele contém informações específicas apresentadas em uma ordem fixa, que são atualizadas quatro vezes ao dia (a cada seis horas)

O código TAF contém as seguintes informações na sequência:

- Grupos de identificação;
- Vento à superfície;
- Visibilidade horizontal predominante prevista;



- Tempo significativo previsto;
- Nuvens previstas (ou visibilidade vertical prevista, se for o caso);
- Temperaturas previstas;
- Grupos de mudanças significativas previstas;
- Código do previsor que confeccionou a previsão.

Assim, um exemplo de mensagem pode ser observado a seguir:

TAF SBPA 130530Z 1312/1412 31015G27KT 8000 SHRA BKN025 TX28/1318Z  
TN20/1409Z BECMG 1400/1402 4000 BKN010 TEMPO 1406/1408 TSRA

Cada grupo de informações possui um significado específico, conforme descrito na Tabela 3.2

| <b>Código</b>   | <b>Descrição</b>   |
|-----------------|--|
| TAF             | Código do relatório (Terminal Aerodrome Forecast)            |
| SBPA            | Identificação do aeroporto (Ex.: Porto Alegre)               |
| 130530Z         | Data/hora da emissão (Dia 13 às 0530 UTC)                    |
| 1312/1412       | Período de validade (de 12h do dia 13 até 12h do dia 14 UTC) |
| 31015G27KT      | Vento de 310°, 15 nós, rajadas de 27 nós                     |
| 8000            | Visibilidade predominante (em metros)                        |
| SHRA            | Fenômeno meteorológico significativo (chuva moderada)        |
| BKN025          | Cobertura de nuvens (broken a 2.500 pés)                     |
| TX28/1318Z      | Temperatura máxima de 28°C prevista às 18 UTC do dia 13      |
| TN20/1409Z      | Temperatura mínima de 20°C prevista às 09 UTC do dia 14      |
| BECMG 1400/1402 | Mudança gradual prevista entre 00 e 02 UTC do dia 14         |
| 4000 BKN010     | Após a mudança, visibilidade 4.000m, nuvens a 1.000 pés      |
| TEMPO 1406/1408 | Condições temporárias previstas entre 06 e 08 UTC do dia 14  |
| TSRA            | Tempestade com chuva forte                                   |

Tabela 3.2: Decodificação de uma mensagem TAF

Seus dados podem ser obtidos por meio de uma API criada pela REDEMET, mantida pelo DECEA, somente após a autorização do departamento.



### 3.1.3 Sistema de Tempo Severo Convectivo (STSC)

O Sistema de Tempo Severo Convectivo (STSC, ou também conhecido como TSC), é uma ferramenta destinada à detecção e ao monitoramento de células convectivas em seu estágio mais ativo. Em outras palavras, o sistema informa aos centros meteorológicos e aos demais usuários da Meteorologia Aeronáutica a ocorrência e a localização de áreas de mau tempo que possam impactar as rotas aéreas.

A coleta desses dados é realizada com base no processamento e análise dos dados dos Radares Meteorológicos e do Sistema de Detecção de Descargas Atmosféricas. A base é atualizada em intervalos inferiores a 3 minutos e possui uma resolução espacial de 10km x 10km.

A requisição para a API pode ser complementada com a data desejada, no formato YYYYMMDDHH, como exemplo, 2025081820 seria 18/08/2025 às 20:00, e também se o usuário gostaria de fazer uma animação com esses dados, tipicamente utilizada para demonstrar o movimento das nuvens em mapas. Sua mensagem de retorno é organizada em um formato JSON, incluindo informações como a posição geográfica do centro da célula (latitude e longitude), seu raio (indicando tamanho/abrangência) e o horário de sua detecção.

A seguir, segue um exemplo de JSON retirado do site oficial do DECEA:

```
{
  "status": true,
  "message": 200,
  "data": {
    "info": {
      "tipo": "3",
      "topo": 999,
      "raio": 5560
    },
    "anima": [
      "14:58"
    ],
    "stsc": [
      [
        {
          "la": "-23.05",
          "lo": "-33.45"
```



```
    },  
    {  
      "1a": "-22.95",  
      "1o": "-34.05"  
    },  
    ...  
  ]  
}  
}
```

Nele, é possível ver a forma como é estruturada as informações coletadas pelo sistema, sendo a chave do objeto *"stsc"* a mais interessante, contendo as posições das células, e também o campo *"info"*, para identificar o tipo da célula, e seu raio.

Esses dados podem ser utilizados em conjunto com dados da trajetória de um voo, para considerar a situação climática em que o avião se encontrava em um determinado ponto de sua rota.

### 3.1.4 SIGMET

O SIGMET é uma mensagem relativa à ocorrência ou previsão de determinados fenômenos meteorológicos (gelo, ciclone, trovoadas e turbulência) em rota e de outros fenômenos na atmosfera (cinzas vulcânicas, nuvem radioativa e vulcão) que possam afetar a segurança das operações aéreas, e a evolução desses fenômenos no tempo e espaço. Seus dados são utilizados para alertar pilotos e controladores para condições severas que requerem atenção máxima, desvio de rotas, ou até mesmo um cancelamento do voo. Eles são validados com frequência variada, mas muito comumente a cada 4 horas.

Para acessar os dados da API, a requisição requer que o país seja especificado, mas também possui três parâmetros opcionais, sendo a data de início e fim, no formato YYYYMMDDHHII e o número de registros por página.

Um exemplo de mensagem SIGMET é mostrado a seguir:

```
SBBS SIGMET 9 VALID 161220/161620 SBBS SEV ICE FCST WI S1623 W04848 -  
S1539 W04853 - S1503 W04822 - S1433 W04730 - S1430 W04656 - S1512 W04623  
- S1559 W04705 - S1646 W04748 - S1623 W04848 FL140/210 STNR NC
```

A explicação de sua estrutura é encontrada na 3.3:



| Termo  | Descrição   |
|--|---|
| SBBS   | Código ICAO da FIR (Região de Informação de Voo) Brasília   |
| SIGMET 9   | 9ª mensagem SIGMET emitida para essa FIR no dia   |
| VALID 161220/161620  | Período de validade: dia 16, das 12:20 UTC às 16:20 UTC   |
| SBBS   | FIR Brasília novamente mencionada   |
| SEV ICE  | Fenômeno meteorológico severo: gelo severo  |
| FCST   | Status do fenômeno: previsão (forecast)   |
| WI S1623 W04848 - S1539 W04853 - S1503 W04822 - S1433 W04730 - ... | Área afetada delimitada por um polígono definido pelas coordenadas geográficas (latitude e longitude) |
| FL140/210  | Altitude do fenômeno entre os níveis de voo FL140 (14.000 pés) e FL210 (21.000 pés)                   |
| STNR   | Movimento do fenômeno: está estacionário (sem deslocamento)   |
| NC   | Mudança esperada: nenhuma (fenômeno deve permanecer estável)  |

Tabela 3.3: Decodificação de uma mensagem SIGMET

Seus dados podem ser utilizados para armazenar mais informações contextuais sobre os fenômenos meteorológicos na rota, e de outros fenômenos que outras bases não possuem, como as cinzas vulcânicas ou nuvens radioativas.

### 3.1.5 GAMET

O GAMET é uma previsão aérea, transmitida em uma linguagem clara abreviada, para voos em níveis baixos, dentro de uma região de voo ou subárea dela. Suas previsões em área são separadas com período de validade de 6 horas, iniciando as 0000, 0600, 1200 e 1800 UTC, cobrindo a camada entre o solo e o FL100 (ou FL150 nas regiões montanhosas).

Uma mensagem GAMET é dividida em duas seções: A primeira detalha informações sobre fenômenos perigosos em rota; a segunda contém informações adicionais que requeiram os voos à baixa altura.

Os campos, obrigatórios e opcionais, para realizar uma requisição à API do REDEMET para obter as mensagens GAMET podem ser observadas na Tabela 3.4.



| Nome                      | Requerido | Descrição  | Valor Padrão | Exemplo      |
|---------------------------|-----------|--|--------------|--------------|
| <i>pais</i>               | Sim       | Nome do País. Países disponíveis para busca: Argentina, Bolívia, Brasil, Chile, Colômbia, Equador, Guiana Francesa, Guiana, Paraguai, Peru, Suriname, Uruguai e Venezuela. | Brasil       | Argentina    |
| <i>data<sub>ini</sub></i> | Não       | Data inicial no formato <i>YYYYMMDDHHII</i>  | Data atual   | 202005120000 |
| <i>data<sub>fim</sub></i> | Não       | Data final no formato <i>YYYYMMDDHHII</i>  | Data atual   | 202005120600 |
| <i>page<sub>tam</sub></i> | Não       | Número de registros por página   | 150          | 100          |

Tabela 3.4: Campos para uma requisição à API

Uma mensagem GAMET é dividida em duas seções. A primeira contém informações sobre fenômenos perigosos em rota e a segunda contém informações adicionais que requeriram os voos à baixa altura. Um exemplo de mensagem GAMET é visto a seguir:

```
SBAZ GAMET VALID 120000/120600 SBGL-SBAZ AMAZONICA FIR BLW FL100
SECN I SFC VIS: 1500M RA N OF S04 AND W OF W050 2000M BR N OF S02
AND E OF W050 SIGWX : ISOL TS N OF S04 AND W OF W050 SIG CLD : ISOL
CB 2700/ABV 10000FT AGL N OF S04 AND W OF W050 ISOL TCU 2500/ABV
10000FT AGL N OF S04 AND W OF W050 ISOL TCU 2500/ABV 10000FT AGL N
OF S02 AND E OF W050 03/06 BKN 300/700FT AGL N OF S01 AND W OF W054
AND E OF W055 03/06 BKN 500/900FT AGL S OF S01 AND N OF S07 AND W
OF W045 AND E OF W060 SECN II PSYS : NIL WIND/T : 2000 FT 080/05KT PS24
5000 FT 050/07KT PS18 10000 FT 100/10KT PS010 CLD : SCT CU 2300/6000FT
AGL SCT AC 9000/ABV 10000FT AGL FZLVL : ABV 10000FT AGL MNM QNH :
1012 HPA SEA: T29 HGT 1,3M VA : NIL=
```

A decodificação da mensagem está na Tabela 3.5, mostrando cada campo e seu significado.



| <b>Campo na mensagem</b>   | <b>Significado</b>  |
|--|---|
| SECN I   | Seção I da mensagem GAMET (fenômenos perigosos em rota)                                       |
| SBAZ   | Indicativo de localidade (código ICAO da FIR Amazônica).                                      |
| GAMET  | Identificação da mensagem, usando a abreviatura GAMET.  |
| VALID 120000/120600  | Grupo data-hora indicando o período de validade (12 às 00:00 UTC até 12 às 06:00 UTC).        |
| SBGL-SBAZ  | Órgão que confeccionou a mensagem (SBGL) seguido de hífen e a FIR aplicável (SBAZ).           |
| AMAZONICA FIR BLW FL100  | Nome da FIR (Amazônica), aplicável a voos abaixo de FL100.                                    |
| SFC VIS: 1500M RA N OF S04 AND W OF W050 2000M BR N OF S02 AND E OF W050       | Áreas extensas de visibilidade abaixo de 5000 m, especificando diferentes regiões.            |
| SIGWX : ISOL TS N OF S04 AND W OF W050   | Tempo significativo: trovoadas isoladas em determinada área.                                  |
| SIG CLD : ISOL CB 2700/ABV 10000FT AGL ...                                     | Áreas extensas de nebulosidade convectiva: CB e TCU em diferentes regiões.                    |
| 03/06 BKN 300/700FT AGL ...  | Nebulosidade dispersa (BKN) com altura da base entre 300 e 700 pés AGL em determinadas áreas. |
| SECN II  | Seção II da mensagem GAMET (ventos, temperaturas, sistemas).                                  |
| PSYS : NIL   | Centros de pressão e frentes: nenhum presente.  |
| WIND/T : 2000 FT 080/05KT PS24, 5000 FT 050/07KT PS18, 10000 FT 100/10KT PS010 | Ventos e temperaturas em altitude para 2000, 5000 e 10000 pés.                                |
| CLD : SCT CU 2300/6000FT AGL, SCT AC 9000/ABV 10000FT AGL                      | Nuvens (fora das já descritas em j), com tipo, base e topo.                                   |
| FZLVL : ABV 10000FT AGL  | Nível de congelamento: acima de 10000 pés AGL.  |
| MNM QNH : 1012 HPA   | QNH mínimo previsto durante o período de validade.  |
| SEA: T29 HGT 1,3M  | Estado do mar: temperatura 29°C, altura significativa de onda 1,3 m.                          |
| VA : NIL   | Cinzas vulcânicas: nenhuma.   |

Tabela 3.5: Exemplo de decodificação da mensagem GAMET SBAZ.



Seus dados podem ser utilizados para adicionar um maior contexto para as rotas de avião de voos de baixa altitude.

### 3.1.6 RADAR

O RADAR retorna imagens de eco de radares meteorológicos, com dados de 01/01/2006 até o presente. Existem diversos tipos de radares, cada um com seu nome e tipo. Os nomes de radares são os seguintes:

- PPI (Plane Position Indicator)
- RHI (Range Height Indicator)
- MAXDISPLAY (Max CAPPI)
- CAPPI (Constant Plan Position Indicator)
- VC (Vertical Cut)
- ET (Echo Top)
- EB (Echo Base)
- VAD (Velocity Azimuth Display)
- VVP (Volume Velocity Processing)

E os tipos são os que seguem:

- Letra Z - refletividade
- Letra V - Velocidade radial média
- Letra W - Velocidade espectral (turbulência)

Os nomes radares disponíveis na API da REDEMET são o MAXDISPLAY e o CAPPI. Uma comparação entre os dois está disponível na Tabela 3.6.



| <b>Nome</b> | <b>Dados fornecidos</b>  |
|-------------|--|
| MAXDISPLAY  | Projeções cartesianas nas direções verticais, norte-sul e leste-oeste, dos valores máximos de refletividade, taxa de precipitação, velocidade radial média ou largura espectral, observados durante uma varredura volumétrica. |
| CAPPI       | Projeção em um plano horizontal constante pré-definido, dos dados de refletividade, taxa de precipitação, velocidade radial média ou largura espectral, obtidos através de uma varredura volumétrica.                          |

Tabela 3.6: Comparação entre o MAXDISPLAY e o CAPPI

A API REDEMETS oferece todos os tipos de produtos do radar. Suas letras sempre estarão presentes ao lado do nome do produto. A comparação entre os diferentes tipos pode ser observada na Tabela 3.7.

| <b>Letra</b> | <b>Tipo de dado</b>  |
|--------------|--|
| Letra Z      | Refletividade, ou seja as cores do produto estão relacionadas com a intensidade da formação e é expressa em dBz. Quanto maior for o dBz, maior será a intensidade da formação. |
| Letra V      | Vento, ou seja, as cores do produto estão relacionadas com a intensidade da velocidade de deslocamento da formação e é expressa em m/s.  |
| Letra W      | Varição da velocidade do vento no interior da formação (turbulência) e é expressa em m/s.  |

Tabela 3.7: Comparação entre os tipos de radares REDEMETS

No caso das velocidades mostradas na escala que acompanha o produto de vento, pode-se notar que existem valores positivos e negativos. Essa distinção é usada para mostrar a direção do deslocamento da nebulosidade em relação à estação radar. As áreas de nebulosidade associadas a valores negativos estão se aproximando da estação radar e as áreas associadas a valores positivos estão se afastando da estação radar.



Para uma requisição à API, o usuário precisa escolher entre as seguintes opções de Nomes:

- **maxcappi**: MaxCappi de 400km de raio;
- **10km**: Cappi de 10,400m, com 250km de raio;
- **07km**: Cappi de 7.300m, com 250km de raio;
- **05km**: Cappi de 5.500m, com 250km de raio;
- **03km**: Cappi de 3.100m, com 250km de raio.

O usuário também precisa escolher os radares meteorológicos cujos dados serão obtidos. Os radares disponíveis estão na Tabela 3.8.



| <b>Código</b> | <b>Radar Meteorológico</b>  |
|---------------|-----------------------------|
| <b>al</b>     | Almenara/MG                 |
| <b>be</b>     | Belém/PA                    |
| <b>bv</b>     | Boa Vista/RR                |
| <b>cn</b>     | Canguçu/RS                  |
| <b>cz</b>     | Cruzeiro do Sul/AC          |
| <b>ga</b>     | Gama/DF                     |
| <b>jr</b>     | Jaraguari/MS                |
| <b>mq</b>     | Macapá/AP                   |
| <b>mo</b>     | Maceió/AL                   |
| <b>mn</b>     | Manaus/AM                   |
| <b>mi</b>     | Morro da Igreja/SC          |
| <b>nt</b>     | Natal/RN                    |
| <b>pl</b>     | Petrolina/PE                |
| <b>pc</b>     | Pico do Couto/RJ            |
| <b>pv</b>     | Porto Velho/RO              |
| <b>sv</b>     | Salvador/BA                 |
| <b>sn</b>     | Santarém/PA                 |
| <b>st</b>     | Santa Teresa/MG             |
| <b>sg</b>     | Santiago/RS                 |
| <b>sf</b>     | São Francisco/MG            |
| <b>ua</b>     | São Gabriel da Cachoeira/AM |
| <b>sl</b>     | São Luís/MA                 |
| <b>sr</b>     | São Roque/SP                |
| <b>tt</b>     | Tabatinga/AM                |
| <b>tf</b>     | Tefé/AM                     |
| <b>tm</b>     | Três Marias/MG              |

Tabela 3.8: Radares Meteorológicos Disponíveis REDEMET.

### 3.2 Instituto Nacional de Meteorologia (INMET)

O INMET, órgão do Ministério da Agricultura e Pecuária, tem como missão promover e coordenar atividades para a produção de informações meteorológicas tempestivas e relevantes voltadas à mitigação de riscos e ao desenvolvimento sustentável do setor agropecuário, para a conservação do meio ambiente e à segurança da sociedade brasileira. (INMET, 2025).

Esse instituto disponibiliza dados de 2000 até o último mês concluído de cada estação meteorológica do Brasil. Essas estações coletam as informações indicadas na Tabela



3.9.

Tabela 3.9: Variáveis meteorológicas do INMET

| Variável                                 | Unidade           |
|--|-------------------|
| Precipitação total                       | mm                |
| Pressão atmosférica                      | mB                |
| Radiação global                          | kJ/m <sup>2</sup> |
| Temperatura do ar                        | °C                |
| Temperatura do ponto de orvalho          | °C                |
| Umidade relativa                         | %                 |
| Direção do vento (norte verdadeiro = 0°) | °                 |
| Rajada máxima do vento                   | m/s               |
| Velocidade do vento                      | m/s               |

O sistema oferece duas modalidades de exportação de dados: por períodos de curta duração, limitados a seis meses, ou por um ano completo. Cada arquivo exportado contém os dados de uma única estação meteorológica, incluindo suas observações e informações complementares sobre a própria estação, como sua identificação e localização geográfica.

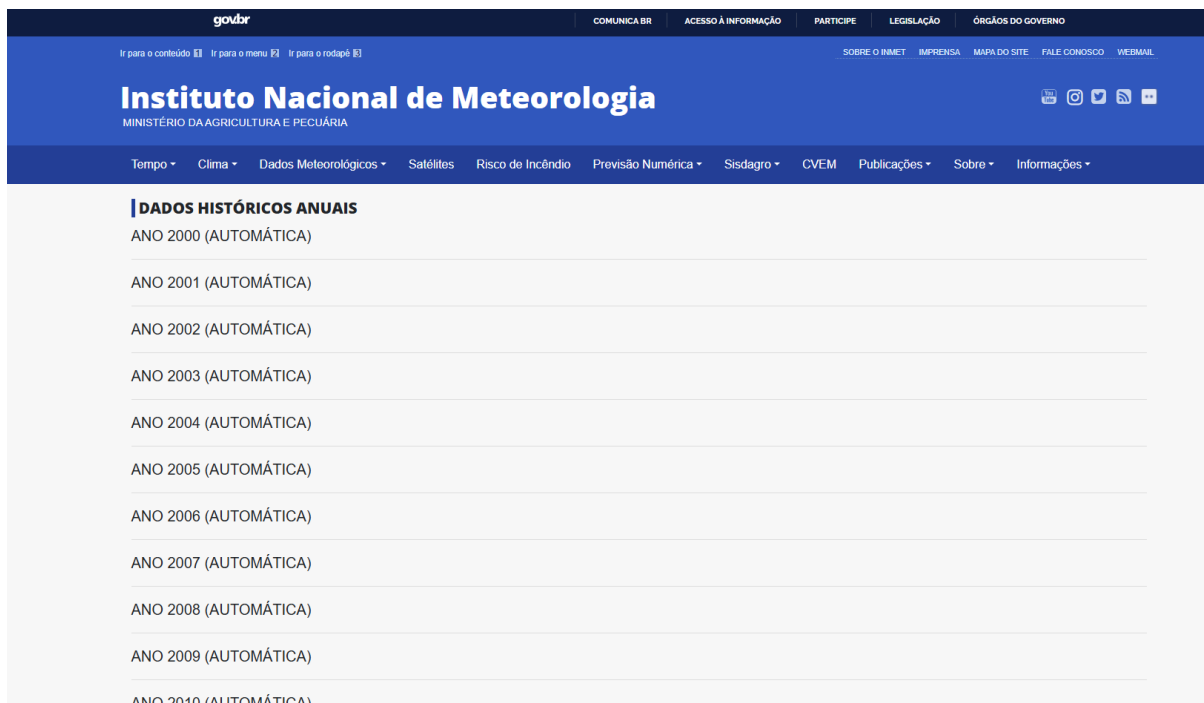


Figura 3.1: Interface de saída do INMET

As estações e suas posições geográficas podem ser visualizadas na Figura 3.2. Observa-se uma concentração significativa de estações meteorológicas nas regiões Su-



deste, Sul e Nordeste do Brasil, especialmente em áreas litorâneas e de maior densidade populacional.

Por outro lado, nota-se uma cobertura reduzida nas regiões Norte e no interior do Centro-Oeste, o que pode limitar a obtenção de dados meteorológicos de alta resolução espacial para operações aeronáuticas nessas áreas.

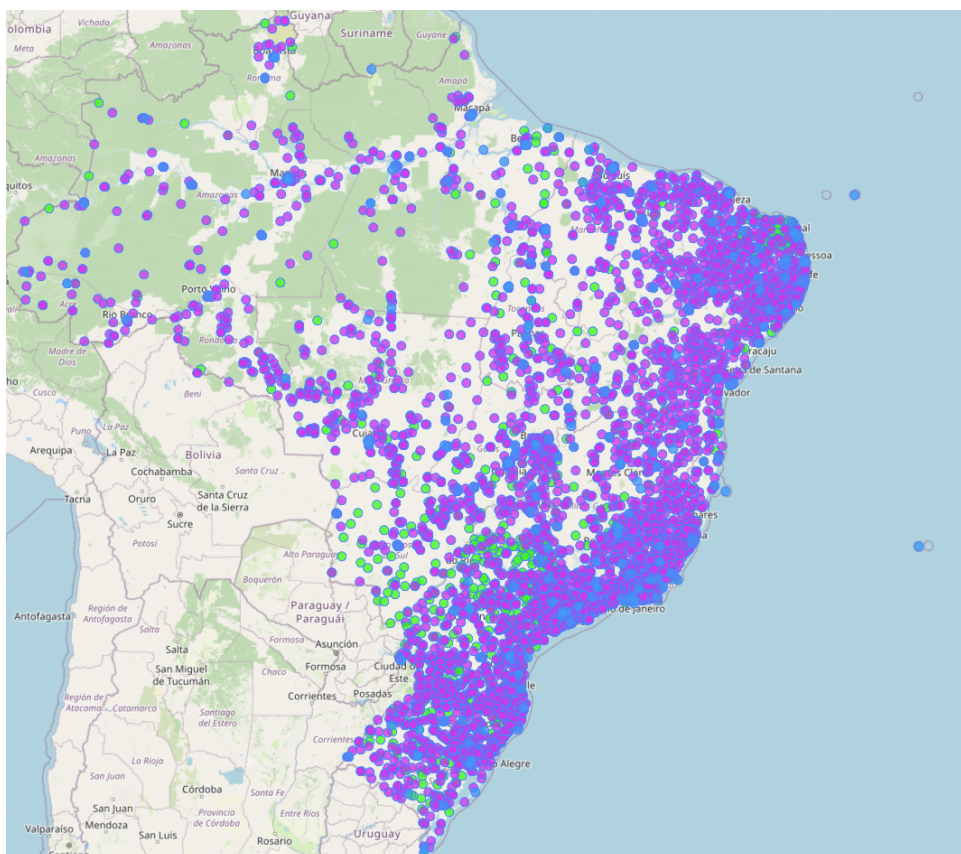


Figura 3.2: Estações meteorológicas do INMET

Em resumo, o INMET constitui uma importante fonte de dados meteorológicos de superfície no Brasil, oferecendo informações consistentes e de boa cobertura temporal desde 2000. Apesar da distribuição desigual das estações — com maior densidade nas regiões mais povoadas e menor presença em áreas remotas —, seus registros são valiosos para análises operacionais, estudos climáticos e validação de outros modelos e bases de dados.

Por fim, no contexto do projeto **AirData**, esses dados podem servir tanto para monitoramento em tempo quase real quanto para complementar séries históricas, especialmente em aeroportos e regiões com observação contínua.



### 3.3 European Analysis 5th Generation (ERA5)

O ERA5 é a quinta geração de reanálise climática global produzida pelo *European Centre for Medium-Range Weather Forecasts* (ECMWF). Ele integra observações meteorológicas de diversas fontes com um modelo atmosférico avançado, por meio de *assimilação de dados*, para gerar um conjunto coerente e completo, com cobertura global.

A cada 12 horas, o sistema ajusta previsões anteriores com novas medições, produzindo uma estimativa otimizada do estado da atmosfera — chamada “análise” — a partir da qual é emitida uma previsão atualizada. Na reanálise, esse mesmo processo é aplicado retrospectivamente, com resolução espacial de cerca de 31 km e temporal horária, permitindo a construção de séries históricas contínuas que se estendem por várias décadas.

Diferentemente da previsão operacional, a reanálise não precisa ser emitida em tempo real. Isso permite a incorporação de um volume maior de medições, incluindo versões corrigidas ou refinadas das originais, o que melhora significativamente a qualidade do produto final.

O ERA5 é atualizado diariamente, com latência de cerca de 5 dias. Esse lançamento preliminar, chamado ERA5T, pode ser revisado caso sejam detectadas falhas graves, resultando em ajustes antes do lançamento final, que ocorre de 2 a 3 meses depois. Nessas situações, os usuários são devidamente notificados. As principais características técnicas do ERA5 estão resumidas na Tabela 3.10.

Tabela 3.10: Resumo técnico do ERA5

| Parâmetro            | Descrição   |
|----------------------|---|
| Produtor             | ECMWF   |
| Tipo                 | Reanálise climática global  |
| Período coberto      | 1940 até presente (atualização diária)  |
| Resolução espacial   | Aproximadamente 31 km   |
| Resolução temporal   | 1 hora  |
| Latência             | Cerca de 5 dias (ERA5T), final 2–3 meses  |
| Principais variáveis | Temperatura, pressão, vento, precipitação, radiação, umidade, etc.  |
| Formatos de dados    | <i>General Regularly-distributed Information in Binary form (GRIB)</i> , <i>Network Common Data Form (NetCDF)</i> , <i>zarr</i> (processados) |

A extração de dados pode ser feita pela plataforma ECMWF *Climate Data Store*, que



permite solicitar informações de 1940 até o dia anterior, escolhendo entre uma ampla variedade de variáveis meteorológicas. A solicitação precisa de autorização para *download*. Em um teste realizado, a equipe do **AirData** solicitou acesso a dados de todas as regiões disponíveis para janeiro de 2025; o pedido foi aceito em 17 minutos.

Outra opção é o serviço *cloud* da Google (Google LLC, 2025), que oferece os dados gratuitamente e sem autorização prévia por meio de um *bucket*, abrangendo o período de 1940 até maio de 2023. Na Tabela 3.11 apresenta-se a comparação entre os métodos de acesso disponíveis.

Tabela 3.11: Comparação entre métodos de acesso ao ERA5

| Critério           | Climate Data Store (CDS)  | Google Cloud  |
|--------------------|---|---|
| Autorização        | Necessária (cadastro e aprovação)   | Não necessária  |
| Período disponível | 1940 até dia anterior à consulta  | 1940 até maio/2023  |
| Custo              | Gratuito  | Gratuito  |
| Formatos           | <i>GRIB, NetCDF, zarr</i>   | <i>GRIB, NetCDF, zarr</i>   |
| Acesso             | Interface web ou API (Python)   | <i>Bucket</i> público no Google Cloud Storage                                       |
| Observações        | Flexível na seleção de variáveis e regiões, porém mais lento para grandes volumes | Ideal para grandes volumes e acesso direto, porém período pode não estar atualizado |

No Google Cloud, os dados são armazenados em três subdiretórios principais, */raw*, */co* e */ar*, conforme ilustra a Figura 3.3.

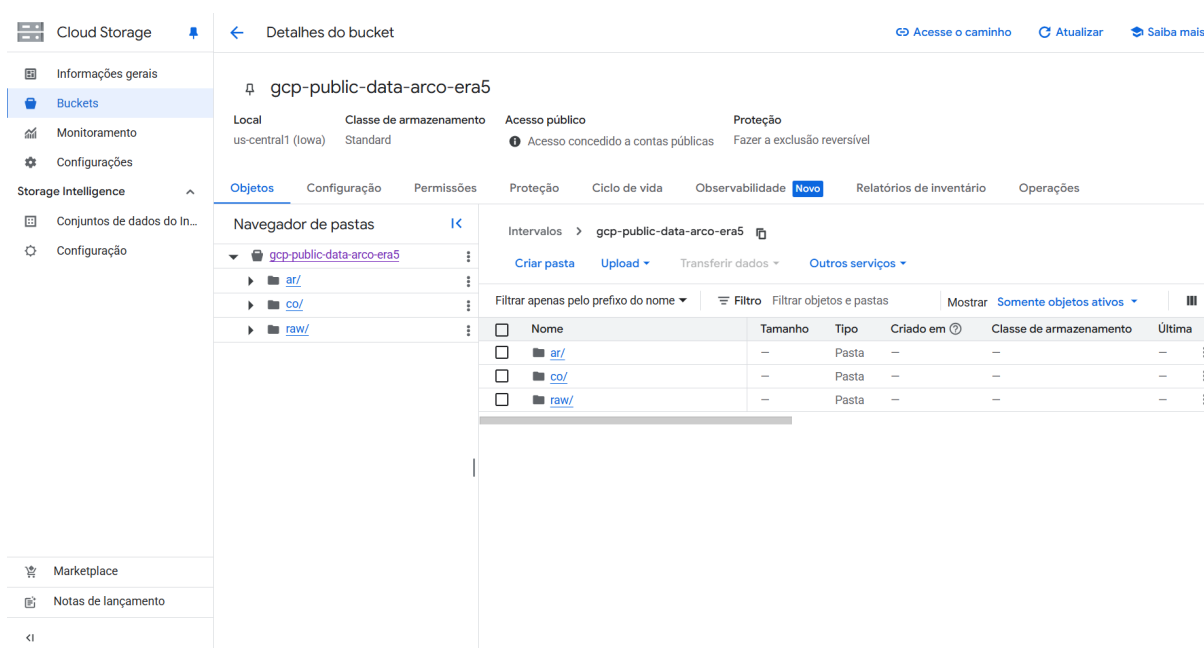


Figura 3.3: Bucket ERA5



No */raw*, são encontrados os dados de origem recebidos do ECMWF. Enquanto no */co*, encontram-se os seguintes dados:

- Umidade no nível do modelo (variáveis relacionadas à umidade e a proporção de mistura de ozônio nos níveis do modelo);
- Vento (divergência, vorticidade, temperatura e velocidade vertical nos níveis do modelo);
- Superfície de nível único: a superfície e o logaritmo de pressão de superfície da superfície do modelo;
- Previsão de nível único: 21 variáveis relacionadas à radiação solar/de onda longa na superfície, quantidade e tipo de precipitação, profundidade de neve e conteúdo de água;
- Reanálise de nível único: 38 variáveis relacionadas à umidade/temperatura do solo, à temperatura e à umidade perto da superfície, ao vapor total de água da coluna e ao condensado, à cobertura total da nuvem e à pressão no mar.

E, por fim, o */ar* contém um conjunto de dados prontos para serem utilizados como *datasets* para algoritmos de machine *learning*. Essa versão está em uma grade regular de latitude/longitude e unifica os dados de superfície e atmosféricos em um único Zarr.

Apesar da ampla cobertura de dados meteorológicos, as restrições de acesso no *Climate Data Store* e a limitação temporal no *Google Cloud* tornam essa base mais adequada como fonte auxiliar, servindo para complementar campos ausentes em outras bases de dado

### 3.4 NCEP-NCAR Reanalysis I e NCEP/DOE Reanalysis II (NCEP)

Nesta seção, encontram-se dois conjuntos de dados globais de reanálise disponíveis: o NCEP-NCAR Reanalysis I e o NCEP/DOE Reanalysis II. Ambos são desenvolvidos pelos Centros Nacionais de Previsão Ambiental dos Estados Unidos (NCEP) (National Centers for Environmental Prediction (NCEP), 2025).

A coleta dos dados, tanto para o Reanalysis I ou II, pode ser feita gratuitamente e sem registros por meio de um pacote na linguagem R chamado *RNCEP* (KEMP et al., 2025), em que pode-se escolher de qual base os dados estão sendo consultados.



### 3.4.1 NCEP-NCAR Reanalysis I

Esta base, desenvolvida em colaboração pelo NCEP e pelo *National Center for Atmospheric Research* (NCAR), cobre o período de 1948 até 6 horas atrás e é continuamente atualizada.

Essa base utiliza-se de diversas observações meteorológicas, coletadas ao longo de décadas e provenientes de múltiplas fontes globais, entre as quais:

- Estações Meteorológicas de Superfície: Fornecem dados sobre pressão ao nível do mar, temperatura, vento e umidade.
- Navios: Medem parâmetros sobre o oceano, como pressão, vento, temperatura da superfície do mar e umidade.
- Aeronaves: Observações durante voos, principalmente de vento e temperatura em diversos níveis da atmosfera.
- Radiossondas (sondas meteorológicas lançadas por balões): Observam perfis verticais de temperatura, vento, umidade e pressão.
- Satélites: Incluem dados de temperatura e vento derivados de sensores remotos desde a década de 1970, além de observações de nuvens por imagens geostacionárias.
- Pibal e Rawindsonde: Equipamentos usados para medir vento em diferentes altitudes (especialmente antes da popularização dos satélites)

Após a coleta dessas informações, os dados são submetidos a processos de controle de qualidade, reduzindo erros ou inconsistências. Enquanto isso, o sistema usado é mantido "congelado" ao longo de todo o período da reanálise, ou seja, o modelo e o sistema de assimilação não são atualizados com o passar do tempo. Isso garante homogeneidade na série temporal.

As observações são, então, combinadas com um modelo numérico global de previsão do tempo, por meio de técnicas de assimilação de dados. A cada 6 horas, o sistema ajusta o modelo com os dados observados. Esse modelo possui uma resolução espacial de 2,5° de latitude por 2,5° de longitude.

Por fim, as variáveis resultantes são classificadas conforme o grau de influência dos dados observacionais:



- A: Fortemente baseadas em observações (alta confiança).
- B: Influenciadas por observações e dependentes do modelo.
- C: Principalmente derivadas do modelo, pouco controladas por observações.
- D: Dados obtidos de fontes externas

Os dados desse projeto, além de serem obtidos pelo pacote RNCEP, podem ser baixados também pela página oficial da *National Oceanic and Atmospheric Administration* (NOAA) - o que pode ser observado na Figura 3.4

The screenshot shows the NOAA Physical Sciences Laboratory website. The page title is "Index of /Datasets/ncep.reanalysis/Monthlies/pressure". It contains a table with columns for Name, Last modified, and Size. The table lists various reanalysis datasets such as 'air.4Xday.1981-2010.ltm.nc', 'air.day.1981-2010.ltm.nc', and 'air.mon.1981-2010.ltm.nc'.

| Name                       | Last modified    | Size |
|----------------------------|------------------|------|
| Parent Directory           |                  | -    |
| air.4Xday.1981-2010.ltm.nc | 2023-02-13 12:46 | 536M |
| air.4Xday.ltm.1981-2010.nc | 2023-02-13 12:46 | 536M |
| air.4Xday.ltm.1991-2020.nc | 2023-02-13 12:46 | 1.0G |
| air.4Xday.ltm.nc           | 2023-02-13 12:46 | 1.0G |
| air.day.1981-2010.ltm.nc   | 2023-02-13 12:46 | 137M |
| air.day.ltm.1981-2010.nc   | 2023-02-13 12:46 | 137M |
| air.day.ltm.1991-2020.nc   | 2023-02-13 12:46 | 260M |
| air.day.ltm.nc             | 2023-02-13 12:46 | 260M |
| air.mon.1981-2010.ltm.nc   | 2023-02-13 12:46 | 5.2M |
| air.mon.inter.std.nc       | 2023-02-13 12:46 | 6.0M |
| air.mon.ltm.1981-2010.nc   | 2023-02-13 12:46 | 5.2M |
| air.mon.ltm.1991-2020.nc   | 2023-02-13 12:46 | 9.3M |
| air.mon.ltm.nc             | 2023-02-13 12:46 | 9.3M |

Figura 3.4: Interface de saída NCEP-NCAR Reanalysis I

### 3.4.2 NCEP/DOE Reanalysis II

Com o uso cada vez mais frequente do Reanalysis I, começaram a ser identificados problemas e anomalias, entre eles:

- Anomalias em dados específicos, resultantes de erros de empacotamento;
- Alta dependência de programas e *scripts* específicos;
- Modelo "congelado" no padrão tecnológico dos anos 1990, com limitações significativas na representação de variáveis, especialmente sobre oceanos e no Hemisfério Sul, onde a cobertura observacional era menor.



Diante disso, o NCEP e o *Department of Energy* (DOE), dos Estados Unidos, uniram recursos e equipes para desenvolver o sucessor, o NCEP/DOE Reanalysis II. Essa nova reanálise incorpora atualizações nas parametrizações físicas (especialmente nos processos de superfície e nos fluxos entre oceano e terra), melhorias no modelo de assimilação de dados e aperfeiçoamentos na representação de variáveis relacionadas à superfície terrestre e aos fluxos de energia.

Além disso, oferece o reprocessamento de arquivos identificados como problemáticos, mantendo as mesmas variáveis, resolução temporal (a cada 6 horas) e resolução espacial ( $2,5^\circ \times 2,5^\circ$ ). O NCEP/DOE Reanalysis II disponibiliza dados desde 1979.

No contexto do **AirData**, o NCEP/DOE Reanalysis II pode ser empregado como fonte complementar para análises atmosféricas históricas e modelagem de cenários, especialmente em regiões e períodos com baixa cobertura observacional direta. Sua resolução espacial de  $2,5^\circ$  e temporal de 6 horas, combinada à consistência metodológica desde 1979, permite preencher lacunas, validar dados de outras fontes e fornecer variáveis meteorológicas em áreas remotas, como rotas oceânicas ou regiões continentais pouco monitoradas.

### 3.5 Considerações Finais

As bases de dados meteorológicas apresentadas neste capítulo — METAR, TAF, STSC, SIGMET, GAMET, RADAR, ERA5, NCEP–NCAR Reanalysis I, NCEP/DOE Reanalysis II e INMET — oferecem diferentes perspectivas sobre as condições atmosféricas, variando em termos de cobertura temporal e espacial, periodicidade de atualização, resolução e confiabilidade.

Enquanto METAR e TAF fornecem informações pontuais e previsões focadas em aeródromos, essenciais para operações aéreas em tempo quase real, STSC, SIGMET, GAMET e RADAR oferecendo previsões focadas no espaço aéreo, as reanálises globais (ERA5, NCEP/NCAR e NCEP/DOE) oferecem séries históricas homogêneas e cobertura em regiões remotas, úteis para análises estratégicas e modelagem de cenários. Por outro lado, o INMET contribui com dados de superfície detalhados para todo o território nacional, fortalecendo o cruzamento e a validação de informações.

No contexto do **AirData**, a integração dessas fontes permitirá construir uma base de conhecimento robusta, combinando observações locais e modelagem global para melhorar a acurácia, preencher lacunas e oferecer suporte tanto a decisões operacionais quanto a estudos de longo prazo. A Tabela 3.12 apresenta um resumo compara-



tivo da aplicabilidade das principais bases meteorológicas discutidas, considerando diferentes fases do voo.

Tabela 3.12: Comparação da aplicabilidade das bases meteorológicas por fase do voo (versão transposta)

| Fonte  | Planejamento pré-voos   | Decolagem   | Subida  | Cruzeiro  | Descida  | Pouso   | Análise pós-voos  |
|--------|---|---|---|---|--|---|---|
| METAR  | Condições atuais no aeródromo de origem/destino                       | Dados locais precisos e oficiais (quando reportados)          | Pouca informação além do aeródromo                  | Não cobre regiões afastadas de aeródromos               | Apenas se próximo ao aeródromo de destino              | Condições reais no momento do pouso (quando reportados)   | Registro oficial das condições no aeródromo               |
| TAF    | Previsão das condições no período do voo                              | Pode indicar tendência, mas sem detalhamento do momento exato | Tendências gerais de evolução das condições         | Não cobre espaço aéreo em rota                          | Previsão específica para a chegada ao aeródromo        | Previsão pode não refletir mudanças repentinas            | Menos útil para análises retroativas                      |
| STSC   | Identificação de áreas convectivas severas previstas ao longo da rota | Pode indicar células próximas à área de decolagem             | Deteção de formações convectivas em desenvolvimento | Alerta sobre presença de tempestades severas na rota    | Localização de células convectivas próximas ao destino | Auxilia na avaliação de risco de aproximação em mau tempo | Permite analisar a presença de tempestades durante o voo  |
| SIGMET | Disponível antes do voo para avaliar fenômenos severos previstos      | Fundamental para decidir viabilidade da decolagem             | Auxilia no planejamento de desvios durante a subida | Indica tempestades, turbulência, gelo ou cinzas em rota | Identifica necessidade de ajustes de aproximação       | Apoia decisões de alternar pouso ou atrasar               | Registro oficial de fenômenos severos enfrentados na rota |

Continuação na próxima página



| <b>Fonte</b> | <b>Planejamento pré-vo</b>  | <b>Decolagem</b>  | <b>Subida</b>   | <b>Cruzeiro</b>  | <b>Descida</b>   | <b>Pouso</b>  | <b>Análise pós-vo</b>   |
|--------------|---|---|---|--|--|---|---|
| GAMET        | Usado no planejamento para voos em baixa altitude                         | Informa riscos meteorológicos locais até FL100/150                | Detecção de turbulência, nevoeiro e CB em rota baixa    | Pouco aplicável ao cruzeiro acima de FL150                                 | Identificação de riscos meteorológicos na aproximação          | Fenômenos relevantes para pousos em baixa altitude                  | Complementa análises de condições em baixa altitude                     |
| RADAR        | Permite observar áreas de precipitação e tempestades antes do vo          | Detecção de células próximas ao aeródromo no momento da decolagem | Visualização de sistemas convectivos no perfil vertical | Acompanhamento em tempo real da evolução de sistemas ao longo da rota      | Localização detalhada de áreas de mau tempo antes da chegada   | Identificação da intensidade de precipitação na final               | Registro da posição de formações meteorológicas para análise retroativa |
| ERA5         | Cenário meteorológico global, inclusive em regiões remotas                | Condições aproximadas na grade (menor precisão local)             | Dados atmosféricos por altitude ao longo de toda a rota | Dados contínuos de vento, temperatura e pressão em qualquer ponto do globo | Condições gerais da região de aproximação (em escala de grade) | Dados aproximados na resolução da grade                             | Série histórica homogênea global para toda a rota                       |
| INMET        | Condições atuais de estações próximas; previsões menos detalhadas que TAF | Dados locais se houver estação no aeródromo ou nas imediações     | Não fornece perfis verticais; foco em superfície        | Sem cobertura em rota (apenas onde existem estações de superfície)         | Útil se houver estação próxima ao destino                      | Condições atuais em superfície (dependente de estação no aeródromo) | Série histórica detalhada de superfície para o território nacional      |





## 4 Bases de Informações Aeronáuticas

Este capítulo detalha as seguintes bases de dados avaliadas como fontes de Informações Aeronáuticas no Brasil:

- AISWEB;
- Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA).

### 4.1 Serviço de Informação Aeronáutica (AISWEB)

O AISWEB é a principal plataforma no Brasil para o acesso a informações aeronáuticas, sendo mantido e operado pelo DECEA, por meio de sua Divisão de Informações Aeronáuticas. Ele funciona como o portal oficial para a divulgação e consulta de dados essenciais para o planejamento e execução de voos no espaço aéreo brasileiro, abrangendo tanto as fases pré-voo quanto as em voo.

A sua importância reside na padronização e na centralização de informações, garantindo que pilotos, controladores de tráfego aéreo, operadores de aeronaves e demais membros da comunidade da aviação tenham acesso a dados atualizados e confiáveis. Para tanto, o AISWEB disponibiliza diversas informações aeronáuticas. As principais podem ser classificadas e detalhadas da seguinte forma:

- **Publicação de Informação Aeronáutica (AIP):** O AIP é o documento oficial que contém informações aeronáuticas permanentes essenciais para a navegação aérea. Ele é composto por três partes principais:
  - *GEN (Generalidades):* Contém informações gerais sobre regras, serviços e procedimentos. Inclui regulamentações nacionais, serviços de tráfego aéreo, serviços de meteorologia, serviços de informação aeronáutica, cartas aeronáuticas, entre outros.
  - *ENR (Em Rota):* Descreve o espaço aéreo brasileiro, as rotas aéreas (aerovias), auxílios à navegação (VOR, NDB, ILS), limites de áreas restritas, perigosas e proibidas, além de procedimentos de navegação em rota.



- *AD (Aeródromos)*: Fornece informações detalhadas sobre aeródromos, incluindo especificações de pistas, elevação, coordenadas, iluminação, facilidades (combustível, manutenção), procedimentos de chegada e saída, informações de contato e características físicas.
- **Suplementos AIP (SUP)**: Publicações temporárias que introduzem alterações de longa duração ao AIP ou fornecem informações temporárias de grande extensão. Podem incluir, por exemplo, a ativação de novas rotas, o fechamento temporário de aeródromos ou a implementação de novos procedimentos de tráfego aéreo.
- **Notice to Airmen (NOTAM)**: Avisos distribuídos por telecomunicações que contêm informações sobre o estabelecimento, condição ou mudança de qualquer facilidade aeronáutica, serviço, procedimento ou perigo, cuja condição é essencial para o pessoal envolvido em operações de voo. Os NOTAMs são importantes para a segurança das operações, pois informam sobre eventos recentes e temporários, como interdição de pistas, falha de auxílios à navegação, atividades de paraquedismo, eventos aéreos, entre outros. O AISWEB permite a consulta de NOTAMs ativos por aeródromo, FIR (Região de Informação de Voo) ou rota específica.
- **Cartas Aeronáuticas**: O AISWEB oferece acesso a uma vasta coleção de cartas aeronáuticas digitais, incluindo:
  - *Cartas de Rota (ENRC)*: Para navegação em rota.
  - *Cartas de Área Terminal (ARC)*: Para operações em torno de grandes aeroportos.
  - *Cartas de Aproximação por Instrumentos (IAC)*: Detalham procedimentos de aproximação para pistas.
  - *Cartas de Saída Padrão por Instrumentos (SID)*: Rotas de saída padronizadas após a decolagem.
  - *Cartas de Chegada Padrão por Instrumentos (STAR)*: Rotas de chegada padronizadas antes da aterrissagem.
  - *Cartas de Aeródromo (ADC)*: Mostram o layout do aeródromo.

Essas cartas são fundamentais para o planejamento de voo e para a navegação visual e por instrumentos.

- **ROTAER**: Um guia auxiliar que contém informações sucintas sobre aeródromos e helipontos, incluindo horários de funcionamento, tipos de tráfego permitido,



facilidades disponíveis e informações de contato. É uma ferramenta de consulta rápida para pilotos.

- **Informações de Frequências de Comunicação:** Detalhes sobre as frequências de rádio utilizadas para comunicação entre aeronaves e controle de tráfego aéreo, incluindo controle de torre, aproximação, área e informações de voo.

O acesso ao AISWEB é geralmente público e gratuito para a consulta da maioria das informações. Sua interface é web-based, permitindo acesso de qualquer dispositivo com conexão à internet. A plataforma é constantemente atualizada para refletir as mudanças no espaço aéreo e nos procedimentos aeronáuticos, garantindo a validade e a precisão das informações fornecidas. A confiabilidade do AISWEB é crucial para a segurança operacional, pois qualquer inconsistência ou desatualização pode comprometer a tomada de decisão em voo.

#### 4.1.1 *Transição de AIS para AIM com base no AIRM*

Historicamente, o Serviço de Informação Aeronáutica (AIS) focava na entrega de produtos, como cartas e publicações impressas. O crescimento da complexidade do tráfego aéreo e a evolução tecnológica impulsionaram a transição para o Gerenciamento da Informação Aeronáutica (AIM), um modelo centrado em dados e orientado à automação, digitalização e interoperabilidade.

Essa transformação é estruturada principalmente sobre o *Aeronautical Information Reference Model (AIRM)*, que fornece a base conceitual e a semântica comum para descrever qualquer tipo de informação aeronáutica. O AIRM atua como a *gramática* da linguagem de informação aeronáutica, garantindo que diferentes sistemas interpretem os dados de forma consistente e interoperável.

A partir do AIRM, derivam-se modelos específicos que definem padrões de dados para diferentes domínios da aviação, permitindo a implementação prática dessa visão. A Tabela 4.1 apresenta os principais modelos e suas aplicações.



Tabela 4.1: Modelos específicos derivados do AIRM e suas aplicações

| Modelo | Aplicação                        | Analogia  |
|--------|----------------------------------|---|
| AIXM   | Infraestrutura e navegação aérea | Dicionário técnico de pistas, rotas e auxílios à navegação        |
| FIXM   | Operações e planos de voo        | Dicionário técnico de procedimentos e voos                        |
| WXXM   | Meteorologia aeronáutica         | Dicionário técnico de boletins e dados meteorológicos             |
| AMXM   | Gerenciamento do espaço aéreo    | Dicionário técnico de setores e restrições de uso do espaço aéreo |

No Brasil, o DECEA conduz essa transição seguindo as diretrizes da ICAO (Anexo 15), adotando padrões internacionais para coleta, validação e disseminação de dados. Essa abordagem garante integridade e rastreabilidade ao longo de todo o ciclo de vida da informação e fortalece a harmonização e interoperabilidade — pilares do projeto **AirData**.

## 4.2 Banco de Dados de Ocorrências Aeronáuticas

Além das informações operacionais e regulatórias disponibilizadas pelo AISWEB, a avaliação da segurança operacional no Brasil é amplamente apoiada pelo banco de dados de ocorrências aeronáuticas do CENIPA.

O CENIPA, organização militar vinculada ao Comando da Aeronáutica, é responsável pela investigação de acidentes e incidentes aeronáuticos com o objetivo primordial de prevenir novas ocorrências, sem atribuir culpa ou responsabilidade. Seu banco de dados consolida informações detalhadas que permitem identificar padrões, avaliar riscos e propor medidas preventivas.

As ocorrências registradas são categorizadas em diferentes tipos, conforme apresentado na Tabela 4.2, o que facilita a organização, a análise e a comparação entre eventos.



Tabela 4.2: Classificação das ocorrências aeronáuticas registradas pelo CENIPA

| <b>Tipo de ocorrência</b> | <b>Descrição</b>  |
|---------------------------|---|
| Acidentes Aeronáuticos    | Eventos em que uma aeronave sofreu danos graves, resultando em fatalidades ou ferimentos graves, ou em que a aeronave está desaparecida ou inacessível.   |
| Incidentes Aeronáuticos   | Ocorrências não classificadas como acidentes, mas que afetam ou poderiam afetar a segurança da operação. Incluem desde falhas mecânicas menores até problemas de comunicação ou infrações de regras de voo. |
| Incidentes Graves         | Incidentes em que as circunstâncias indicam que um acidente quase ocorreu.  |

Para cada evento registrado, o banco de dados armazena um conjunto estruturado de campos que permitem a análise detalhada do ocorrido. Esses campos, exemplificados na Tabela 4.3, abrangem desde dados básicos, como data e localização, até informações analíticas, como fatores contribuintes e recomendações de segurança.

Tabela 4.3: Exemplos de campos registrados pelo CENIPA para cada ocorrência

| <b>Campo</b>                    | <b>Descrição</b>  |
|---------------------------------|---|
| Data e Hora da Ocorrência       | Permite a análise temporal dos eventos.   |
| Localização                     | Coordenadas geográficas ou referência do aeródromo/área.  |
| Tipo de Aeronave                | Modelo, prefixo e informações sobre a aeronave envolvida.   |
| Fase do Voo                     | Fase da operação (decolagem, cruzeiro, aproximação, pouso) na qual ocorreu o evento.                                  |
| Classificação da Ocorrência     | Indica se foi acidente, incidente grave ou incidente.   |
| Fatores Contribuintes           | Elementos que contribuíram para a ocorrência, podendo ser humanos, materiais, ambientais, operacionais ou gerenciais. |
| Recomendações de Segurança      | Medidas propostas para prevenir a recorrência de eventos semelhantes.   |
| Relatório Final da Investigação | Documento com a descrição, análise e conclusões sobre a ocorrência.   |
| Vítimas e Danos                 | Informações sobre fatalidades, feridos e extensão dos danos à aeronave ou a terceiros.                                |

O acesso a essa base de dados é público e pode ser feito por meio do site oficial do CENIPA, geralmente com ferramentas de busca e relatórios estatísticos. Embora seja utilizada principalmente para fins de prevenção, a base também é um recurso valioso para pesquisas acadêmicas, análises estatísticas e desenvolvimento de modelos preditivos de segurança de voo no Brasil.



### 4.3 Considerações Finais

As bases de informações aeronáuticas analisadas neste capítulo - AISWEB e CENIPA - oferecem contribuições complementares para o contexto operacional e estratégico da aviação no Brasil.

O AISWEB centraliza e padroniza informações essenciais para a condução segura das operações aéreas, fornecendo dados atualizados que abrangem desde o planejamento até a execução do voo. Por outro lado, o banco de dados do CENIPA reúne registros detalhados de ocorrências aeronáuticas, permitindo a identificação de padrões de risco, a avaliação da eficácia de medidas preventivas e o apoio a pesquisas e análises estatísticas.

A integração dessas fontes no contexto do **AirData** pode potencializar a capacidade de monitoramento e análise, combinando dados operacionais em tempo quase real com informações históricas e investigativas, fortalecendo assim a segurança e a eficiência das operações aéreas.

Tabela 4.4: Comparação entre AISWEB e CENIPA

| Aspecto                 | AISWEB  | CENIPA  |
|-------------------------|---|---|
| Finalidade              | Disponibilizar informações aeronáuticas para planejamento e execução de voos          | Investigar e registrar ocorrências aeronáuticas para fins de prevenção                            |
| Tipo de Informação      | Dados operacionais e regulatórios (AIP, SUP, NOTAM, cartas, frequências)              | Registros de acidentes, incidentes e incidentes graves, com fatores contribuintes e recomendações |
| Periodicidade           | Atualização contínua, refletindo mudanças operacionais e regulatórias                 | Atualização conforme registro e conclusão de investigações  |
| Cobertura Temporal      | Foco no presente e no curto prazo (operações em andamento e próximas)                 | Histórico de ocorrências ao longo de anos   |
| Usuários Principais     | Pilotos, controladores de tráfego aéreo, operadores de aeronaves e órgãos reguladores | Investigadores, analistas de segurança, gestores e pesquisadores                                  |
| Papel no <b>AirData</b> | Fonte primária de dados operacionais em tempo quase real                              | Base histórica e investigativa para análise de segurança e modelagem de risco                     |



## 5 Bases Normativas

Este capítulo descreve, de forma abreviada, o arcabouço normativo que regula as operações de transporte aéreo no Brasil e no cenário internacional, bem como como essas normas serão incorporadas à base de conhecimento do Projeto **AirData**.

O Código Brasileiro de Aeronáutica (Lei nº 7.565/1986) (BRASIL, 1986) estabelece os princípios gerais da aviação civil no país, incluindo a disciplina do espaço aéreo, requisitos de aeronavegabilidade, diretrizes para a investigação de acidentes e a aplicação de sanções. A Lei do Aeronauta (Lei nº 7.183/1984) (BRASIL, 1984) define direitos e deveres dos tripulantes, com ênfase em jornadas de trabalho, períodos de descanso e critérios de remuneração.

No âmbito regulatório, a ANAC emite os Regulamentos Brasileiros da Aviação Civil (RBAC) e Regulamentos Brasileiros de Homologação Aeronáutica (RBHA), que têm força de norma e tratam de temas como operações de voo, manutenção, certificação, licenças e requisitos de segurança. As Instruções Suplementares (IS) complementam os RBAC e RBHA, apresentando detalhamentos técnicos e exigências específicas.

Além disso, o arcabouço normativo é complementado pelas Resoluções da ANAC, que são atos normativos que introduzem mudanças regulatórias de caráter obrigatório, frequentemente incluindo emendas técnicas aos RBAC, definição de requisitos operacionais específicos ou regulamentos gerais de transporte aéreo. São especialmente relevantes ao AirData porque fornecem condições de operação (como segurança, performance operacional e direitos dos passageiros) que influenciam diretamente indicadores como pontualidade, atrasos e conformidade regulatória.

Complementarmente, as Portarias da ANAC desempenham um papel administrativo normativo, definindo prazos, procedimentos, programas e regulamentações secundárias que operacionalizam o arcabouço regulatório. No contexto do AirData, elas são úteis para rastrear agendas regulatórias, implementar regras administrativas temporárias ou específicas e direcionar mudanças operacionais de curto prazo.

O DECEA controla a atualização temporal das informações aeronáuticas por meio do calendário AIRAC, ciclo de 28 dias que define a vigência de cartas, procedimentos e demais publicações do AIS disponibilizadas via AISWEB. O DECEA também publica as Instruções do Comando da Aeronáutica (ICA), estabelecendo procedimentos ope-



racionais, regras de serviço de tráfego aéreo e protocolos de resposta a incidentes.

No plano internacional, a ICAO publica os Anexos da Convenção de Chicago, que contêm normas e práticas recomendadas da *Standards and Recommended Practices* (SARP) adotadas como referência global. Esses Anexos, assim como os *Procedures for Air Navigation Services* (PANS) e demais documentos técnicos da ICAO, são incorporados ao e-AIP brasileiro e sincronizados com os ciclos AIRAC pertinentes.

Para integrar esse conjunto normativo ao repositório semântico do Projeto **AirData**, cada documento será capturado e representado com um conjunto fixo de atributos, conforme a Tabela 5.1, garantindo rastreabilidade e permitindo vinculação a eventos operacionais.

Tabela 5.1: Atributos previstos para representação de normas no repositório do **AirData**

| Atributo            | Descrição  |
|---------------------|--|
| Tipo de norma       | Lei, regulamento, instrução, manual internacional.                           |
| Identificação       | Número, título e versão do documento.  |
| Datas relevantes    | Data de publicação e data de vigência AIRAC.                                 |
| Escopo geográfico   | Abrangência: nacional, regional ou internacional.                            |
| Escopo operacional  | Área de aplicação: operações de voo, controle de tráfego, certificação, etc. |
| Fonte de origem     | Órgão emissor, como ANAC, DECEA ou ICAO.                                     |
| Identificador único | Código interno para controle e rastreamento de versões.                      |

As versões históricas serão preservadas para permitir análises temporais das alterações normativas e comparações entre estados anteriores e vigentes.

A captação e processamento desses dados normativos combinarão mecanismos automáticos e estruturados. Sempre que disponível, serão utilizados pontos de acesso programáticos (APIs) para capturar textos normativos; nos casos em que não houver interface estruturada, será aplicado *web scraping* controlado dos portais aplicáveis.

Após a captura, os dados serão relacionados a uma ontologia aeronáutica a ser desenvolvida no **AirData**. Nesse contexto, a ontologia será um modelo conceitual formal que organizará e descreverá, de maneira hierárquica e padronizada, os termos, entidades e relações do domínio da aviação. Essa estrutura permitirá que os textos normativos sejam interpretados de forma consistente por sistemas automatizados, viabilizando análises complexas.

De forma sumária, o fluxo de processamento previsto incluirá:



1. Captura do documento bruto e extração dos metadados relevantes;
2. Normalização e padronização dos campos (identificadores, datas, tipos);
3. Atribuição de identificador único e registro de versão;
4. Enriquecimento semântico, associando o conteúdo a conceitos definidos na ontologia aeronáutica do projeto;
5. Armazenamento incremental com histórico imutável para permitir a reconstrução de estados normativos em datas específicas.

Em fases futuras, cada norma será decomposta para extração de regras operacionais explícitas e implícitas. O sistema permitirá comparar versões e identificar, por exemplo, quando uma mudança normativa entrar em vigor e quais operações de voo ocorreram sob diferentes regimes. Essa correlação permitirá validações como: verificar se uma aeronave operou conforme os requisitos vigentes em uma data específica ou avaliar impactos normativos sobre indicadores operacionais.

O objetivo técnico dessa integração será produzir pipelines ETL capazes de incorporar a norma como uma camada de restrição semântica. Esses pipelines deverão permitir consultas combinadas entre dados operacionais, geográficos e normativos, respondendo a questões como: “Qual será o requisito vigente para determinado procedimento de voo na data X?” ou “Haverá mudança normativa que coincida com variação observada em um indicador de atraso?”. Portanto, a estrutura resultante deverá suportar raciocínio temporal e condicionamento de análises com base na vigência normativa.

Além do conjunto normativo já identificado, serão avaliadas outras normas e documentos técnicos a partir da validação do escopo principal de uso do **AirData** junto à ANAC e ao DECEA. Por exemplo, caso o foco inicial do sistema seja a análise de atrasos operacionais, serão priorizados, em um primeiro momento, os regulamentos e publicações diretamente relacionados a esse tema. Em fases subsequentes, o repositório normativo poderá ser expandido para contemplar outros tópicos, assegurando que a base de conhecimento acompanhe as necessidades operacionais e estratégicas do projeto.

## 5.1 Considerações Finais

O levantamento e a organização do arcabouço normativo são elementos essenciais para que o **AirData** possa, no futuro, responder a consultas complexas com base em



interpretações semanticamente consistentes das regras vigentes. Nesta fase de concepção, o objetivo central é identificar a relevância das normas para a estruturação do projeto e estabelecer o modelo de representação que assegurará rastreabilidade, controle de versões e capacidade de análise temporal.

A seleção inicial contempla o conjunto normativo de maior impacto para o escopo validado, mas a definição de demais normas e documentos técnicos a serem incorporados será feita em conjunto com os stakeholders institucionais — ANAC, SAC e DECEA —, garantindo alinhamento às prioridades operacionais e estratégicas. Por exemplo, caso o foco inicial do sistema seja a análise de atrasos operacionais, serão priorizados, em um primeiro momento, os regulamentos e publicações diretamente relacionados a esse tema.

A Tabela 5.2 apresenta uma síntese dos principais conjuntos normativos considerados, destacando sua origem, abrangência e papel no contexto do projeto.

Tabela 5.2: Comparação dos principais conjuntos normativos no contexto do **AirData**

| Norma                            | Órgão emissor      | Abrangência   | Papel no projeto  |
|----------------------------------|--------------------|---------------|---|
| Código Brasileiro de Aeronáutica | Congresso Nacional | Nacional      | Define princípios gerais da aviação civil e disciplina do espaço aéreo.   |
| Lei do Aeronauta                 | Congresso Nacional | Nacional      | Regula direitos, deveres e condições de trabalho dos tripulantes.   |
| RBAC/RBHA                        | ANAC               | Nacional      | Estabelece requisitos técnicos e operacionais para segurança de voo.  |
| IS                               | ANAC               | Nacional      | Complementa RBAC/RBHA com detalhamento técnico.   |
| Resoluções                       | ANAC               | Nacional      | Estabelecem direitos obrigatórios (ex.: direitos de passageiros, emendas a regulamentos técnicos), com caráter vinculante e abrangente. |
| Portarias                        | ANAC               | Nacional      | Instrumentos administrativos para definir prazos, procedimentos, programas e regras específicas de aplicação normativa.                 |
| ICA                              | DECEA              | Nacional      | Define procedimentos operacionais e de tráfego aéreo.   |
| Anexos e PANS da ICAO            | ICAO               | Internacional | Normas e práticas recomendadas (SARPs) para harmonização global.  |



## 6 Dados Abertos

O projeto **AirData** será fundamentado nos princípios de dados abertos e nas diretrizes estabelecidas pela legislação brasileira que regulamenta o tema, incluindo a Lei de Acesso à Informação (Lei nº 12.527/2011) (BRASIL, 2011) e a Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018) (BRASIL, 2018).

Essa abordagem irá assegurar que as informações tratadas no âmbito do projeto sejam disponibilizadas de forma transparente, acessível e interoperável, respeitando simultaneamente as restrições legais de privacidade e segurança.

Além de cumprir com os requisitos normativos, a adoção dos conceitos de dados abertos será conectada diretamente à arquitetura técnica e às operações do sistema. Assim, desde a fase de concepção, o modelo de dados e os fluxos de integração considerarão padrões e formatos abertos, viabilizando consultas, análises e correlações automáticas entre diferentes fontes.

### 6.1 Conceitos

Evoluções tecnológicas viabilizaram uma nova plataforma denominada dados abertos. Dados abertos são dados disponibilizados em formatos pré-determinados que possibilitam sua utilização e reutilização por terceiros, sendo legíveis por máquinas e compatíveis com dispositivos da Tecnologia da Informação e Comunicação (TIC).

Outro conceito amplamente aceito é o da *Open Knowledge Foundation* (OKF), que define dados abertos como aqueles disponíveis de forma que qualquer pessoa possa usar, reutilizar e redistribuir, com a exigência, no máximo, de atribuição de autoria e compartilhamento sob a mesma licença. Essa condição é geralmente satisfeita pela publicação em formato e licença abertos.

### 6.2 Premissas e Questões Técnicas

Existem premissas básicas para a conformidade com o conceito de dados abertos. As principais são resumidas em três leis e oito princípios.



A Tabela 6.1 apresenta as três leis fundamentais que norteiam o conceito de dados abertos.

Tabela 6.1: As três leis de dados abertos

| Item                     | Significado   |
|--------------------------|---|
| Disponibilidade e acesso | Totalmente disponíveis na Internet, com custo acessível e formato reutilizável. |
| Reuso e redistribuição   | Formato deve permitir reutilização e cruzamento com outros dados.               |
| Participação universal   | Disponível para todos, sem restrições (ex.: apenas uso educacional).            |

Em complemento, a Tabela 6.2 descreve os oito princípios complementares que orientam a disponibilização e utilização de dados abertos.

Tabela 6.2: Oito princípios de dados abertos

| Item                        | Significado   |
|-----------------------------|---|
| Completos                   | Todos os dados públicos devem estar disponíveis.          |
| Primários                   | Dados com granularidade máxima, sem agregações.           |
| Atuais                      | Quanto mais atualizados, maior a utilidade.               |
| Acessíveis                  | Disponíveis sem restrições para os interessados.          |
| Compreensíveis por máquinas | Estruturados em formatos reutilizáveis.                   |
| Não discriminatórios        | Acesso sem procedimentos limitantes (ex.: cadastro pago). |
| Não proprietários           | Sem controle exclusivo por uma organização.               |
| Livres de licenças          | Não submetidos a copyright ou patentes.                   |

Nem todas as informações podem ser divulgadas de forma aberta. Muitas contêm dados pessoais ou de organizações, protegidos por legislações específicas. No Brasil, a Lei Geral de Proteção de Dados Pessoais (LGPD) regulamenta o tratamento de dados, garantindo proteção à liberdade, privacidade e formação da personalidade individual.

Além das exigências legais, destacam-se recomendações técnicas:

- Licenças abertas, como as *Creative Commons*.
- Interoperabilidade entre sistemas e fontes de dados.
- Disponibilidade para download em formato legível por máquina.



Também é recomendável fornecer metadados — descrições que facilitam a compreensão e utilização dos dados — organizados em catálogos de dados.

Berners-Lee (2009) propôs um modelo de maturidade de dados abertos com cinco níveis:

1. Publicação na Web com licença aberta.
2. Dados legíveis por máquina, mesmo em formato proprietário.
3. Dados em formato aberto e legível por máquina.
4. Dados seguindo todos os princípios de dados abertos.
5. Dados interconectados, usando tecnologias de dados ligados.

Outro ponto essencial é a definição clara de políticas de governança de dados, contemplando responsabilidades, procedimentos de atualização, mecanismos de controle de qualidade e estratégias de preservação digital. Essas políticas devem estar alinhadas tanto às normas legais quanto às necessidades dos usuários, garantindo que os dados abertos se mantenham relevantes, confiáveis e tecnicamente compatíveis com novas plataformas e aplicações.

Além disso, a segurança da informação deve ser incorporada desde a concepção do sistema, adotando práticas como anonimização e pseudonimização quando necessário, autenticação robusta para acesso a camadas sensíveis e auditorias periódicas para identificar e corrigir vulnerabilidades. A combinação de transparência, qualidade e segurança cria uma base sólida para que a disponibilização de dados abertos seja sustentável e confiável ao longo do tempo.

### 6.3 Considerações Finais

A adoção das premissas e princípios de dados abertos no **AirData** representa um compromisso com a transparência e a inovação no setor aeronáutico brasileiro. Ao integrar padrões técnicos, requisitos legais e boas práticas de governança, o projeto fortalece a capacidade de gerar análises precisas, fomentar a colaboração entre instituições e oferecer informações de alto valor para a sociedade.

Esse alinhamento entre legislação, tecnologia e gestão de dados permitirá que o **AirData** evolua continuamente, incorporando novas fontes e formatos, mantendo a interoperabilidade e garantindo o equilíbrio entre abertura e proteção da informação.

## 7 Propostas de Integração entre Bases de Dados

O projeto **AirData** visa desenvolver um sistema de inteligência artificial especializado capaz de fornecer análises contextualizadas e suporte inteligente à tomada de decisão no setor de transporte aéreo brasileiro. Para atingir este objetivo, o sistema deve integrar e processar simultaneamente dois tipos fundamentais de dados: analíticos e semânticos; conforme ilustrado na Figura 7.1.

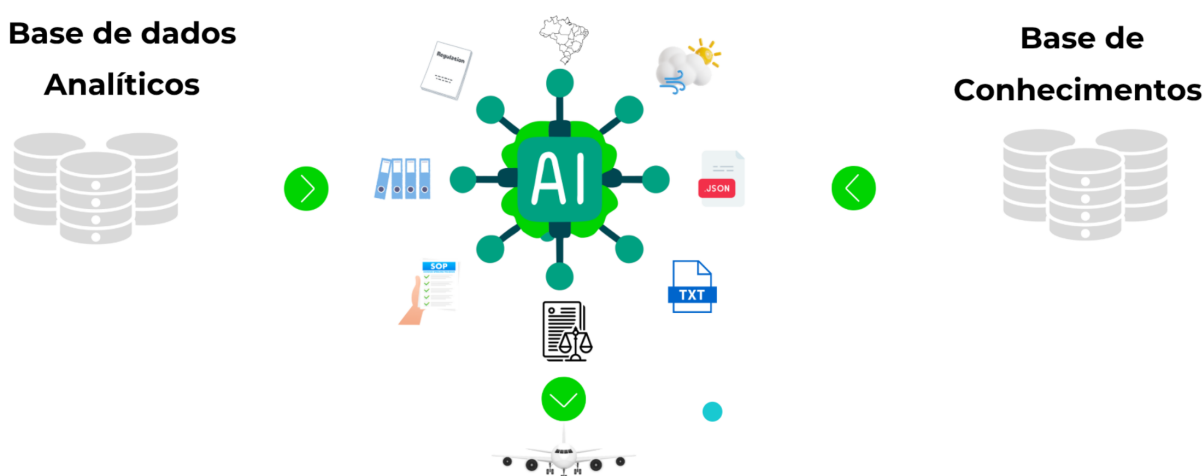


Figura 7.1: Arquitetura Conceitual do Sistema **AirData**

De forma resumida, a Figura 7.1 apresenta a visão geral do desafio do **AirData**: desenvolver uma inteligência artificial que seja capaz de:

- **Processar dados operacionais estruturados** (à esquerda): informações sobre voos, aeroportos, atrasos, rotas e operações;
- **Acessar conhecimento contextual e normativo** (à direita): regulamentações, procedimentos, manuais técnicos e conhecimento especializado do domínio aeronáutico;
- **Integrar ambas as dimensões** para gerar respostas que não apenas apresentem fatos, mas os contextualizem adequadamente segundo as normas e práticas do setor.

Esta integração permite que o sistema responda questões como, por exemplo, "Por que o voo LATAM 5411 atrasou?" de forma completa: identificando os dados relaci-



onais (45 minutos de atraso por condições meteorológicas) e contextualizando-os com o conhecimento regulatório pertinente (Ex. ICA 100-12 sobre procedimentos de Tráfego Aéreo).

Portanto, o desafio central reside em como harmonizar e transformar dados fragmentados, complexos e distribuídos, oriundos de múltiplas fontes com formatos, semânticas e níveis de qualidade distintos, em uma base analítica universal, confiável e interoperável, que sirva de suporte para o desenvolvimento desta inteligência artificial especializada.

Em síntese, a inovação do **AirData** está na integração sinérgica entre as duas bases destacadas por meio de técnicas de IA, permitindo que o sistema não apenas acesse dados e conhecimento separadamente, mas os combine de forma inteligente para fornecer análises contextualizadas que atendam às especificidades do setor aeronáutico brasileiro.

### 7.1 Desafios das Bases de Dados Analíticas

Apesar da crescente disponibilidade de dados aeronáuticos estruturados no Brasil, o processo tradicional de integração e análise dessas informações ainda apresenta limitações importantes.

Conforme ilustrado na Figura 7.2, a arquitetura predominante ainda se baseia em fluxos tradicionais de coleta e consolidação de dados, nos quais informações provenientes de fontes diversas, como: ANAC, DECEA, OpenSky e REDEMET; são extraídas, transformadas e carregadas manualmente ou por meio de rotinas específicas para sistemas centrais de análise.

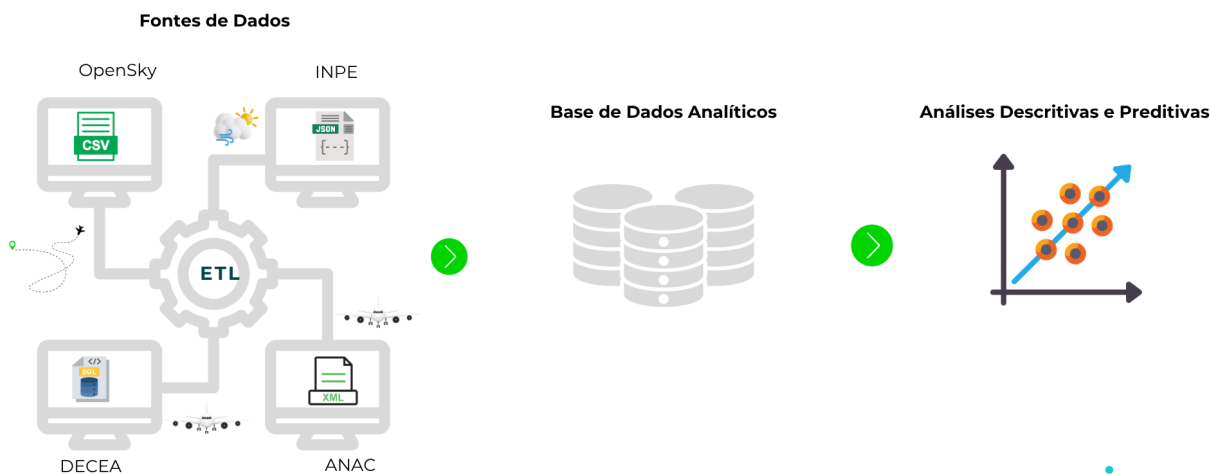


Figura 7.2: Limitações do Fluxo Tradicional de Análise de Dados



Por fim, esses dados são então utilizados em análises pontuais, geralmente de natureza descritiva ou preditiva, mas com baixa flexibilidade e integração semântica entre as fontes.

Ademais, o modelo ilustrado na Figura 7.2, apesar de ser uma prática comum de armazenamento e análise de dados, ainda é marcado por fragmentação institucional, diversidade de formatos, variações na granularidade temporal e, principalmente, pela falta de padronização semântica entre as múltiplas fontes.

Essa falta de alinhamento conceitual compromete a interoperabilidade entre sistemas, dificulta a correlação entre causas e efeitos operacionais e inviabiliza o uso eficiente de técnicas de inteligência artificial.

Como resultado, o conhecimento extraído permanece restrito a análises retrospectivas e isoladas, limitando sua utilidade para a tomada de decisão estratégica e tática no setor.

Resumidamente, as principais limitações desse fluxo tradicional incluem:

- **Integração superficial das fontes:** cada base mantém sua estrutura original sem harmonização semântica;
- **Discrepâncias de granularidade temporal:** diferentes frequências de atualização dificultam correlações precisas;
- **Formatos heterogêneos:** diversidade de formatos (CSV, XML, JSON, SQL) exigem múltiplas rotinas de processamento;
- **Ausência de modelo semântico comum:** falta de padronização conceitual gera ambiguidades;
- **Limitações para raciocínio automatizado:** estrutura inadequada para aplicação de IA avançada.

Essas limitações evidenciam que o modelo tradicional de integração de dados, baseado em pipelines isolados e estruturas analíticas centralizadas e pouco modulares (isto é, sistemas rígidos e difíceis de adaptar a novas fontes de dados), não é capaz de atender às demandas de interoperabilidade, escalabilidade e inteligência exigidas pelo escopo do **AirData**.

Diante disso, torna-se necessário investigar, de forma sistemática, as abordagens mais eficazes já consolidadas na literatura técnica e científica para lidar com ecossistemas de dados distribuídos, heterogêneos e dinâmicos. Essa investigação deve



considerar tanto soluções arquiteturais quanto modelos conceituais e estratégias de integração semântica, com vistas a fundamentar as propostas técnicas apresentadas nos próximos produtos.

## 7.2 Desafios da Base de Conhecimento Semântica

Para completar a arquitetura do **AirData**, além da base analítica, é necessária a construção da base de conhecimento semântica (lado direito da Figura 7.1). Essa base tem como função organizar e representar, de forma estruturada, o conhecimento normativo, técnico e contextual que rege as operações do setor aeronáutico.

Para ilustrar, a Figura 7.3 destaca os principais componentes que integram essa base: desde regulamentos e procedimentos operacionais até documentos técnicos, dados meteorológicos, geográficos e arquivos em diferentes formatos. Esses elementos, embora diversos em natureza e origem, devem ser integrados em uma estrutura semântica coerente e interoperável.



Figura 7.3: Componentes da Base de Conhecimento Semântica

Diferentemente das bases tradicionais, centradas majoritariamente em dados numéricos e estruturados, a base de conhecimento semântica do **AirData** deverá incorporar informações contextuais essenciais para a compreensão e interpretação dos dados operacionais. Isso inclui por exemplo:



- **Conhecimento Regulatório:** normas ANAC (RBACs), instruções DECEA (ICAs) e regulamentações ICAO;
- **Conhecimento Técnico-Operacional:** procedimentos padrão, manuais de operação, especificações técnicas;
- **Conhecimento Contextual:** relações causais, interpretações de códigos, explicações de fenômenos;
- **Conhecimento Geográfico:** informações sobre aeroportos, rotas, espaços aéreos;
- **Conhecimento Meteorológico:** interpretação de condições climáticas e impactos operacionais.

A construção dessa base será imprescindível para que o sistema **AirData** vá além da simples recuperação de dados e passe a fornecer análises contextualizadas e fundamentadas em normas, procedimentos e saberes especializados. No entanto, a diversidade, a complexidade e a natureza não estruturada dessas fontes impõem desafios significativos quanto à sua representação, organização e integração.

Portanto, será necessário identificar modelos conceituais e tecnologias já consolidadas na literatura técnica e científica que possibilitem estruturar esse conhecimento de maneira formal, interpretável por máquinas e interoperável com os dados operacionais. Tal esforço será essencial para viabilizar uma camada semântica robusta, que permita ao sistema compreender, correlacionar e explicar eventos operacionais à luz do conhecimento normativo e técnico do setor aeronáutico.

### 7.3 Ontologias para o AirData

Para superar os desafios identificados e viabilizar a arquitetura conceitual apresentada na Figura 7.1, o projeto **AirData** irá adotar as ontologias como conceito fundamental. Uma ontologia, no contexto da representação de conhecimento, constitui uma especificação formal e explícita de uma conceitualização compartilhada (GRUBER, 1993).

No domínio aeronáutico, as ontologias oferecem capacidades essenciais para harmonizar semanticamente dados de fontes heterogêneas. Elas resolvem ambiguidades terminológicas entre diferentes sistemas e representam conhecimento contextual de forma estruturada.



Além do mais, permitem que sistemas de IA compreendam não apenas dados factuais, mas também suas implicações regulatórias e operacionais, facilitando raciocínio automatizado por meio de relações formalmente definidas entre conceitos do domínio aeronáutico.

O setor aeronáutico atualmente reconhece o valor das ontologias por meio de iniciativas consolidadas que servem como fundamento teórico e prático para o **AirData**. Por exemplo, a *NATIONAL AERONAUTICS AND SPACE ADMINISTRATION* (NASA) desenvolveu uma ontologia para o gerenciamento de tráfego aéreo (KELLER; JOBE; MULFINGER, 2017), que integra dados de múltiplas fontes do sistema aeronáutico americano.

A ontologia da NASA demonstra a viabilidade prática de usar representação semântica para integrar dados operacionais complexos, cobrindo aspectos de voo e navegação, equipamentos de aeronaves, infraestrutura do espaço aéreo, meteorologia e iniciativas de gerenciamento de tráfego aéreo. Como exemplo, a Figura 7.4 destaca uma representação gráfica esquemática da ontologia para um voo.

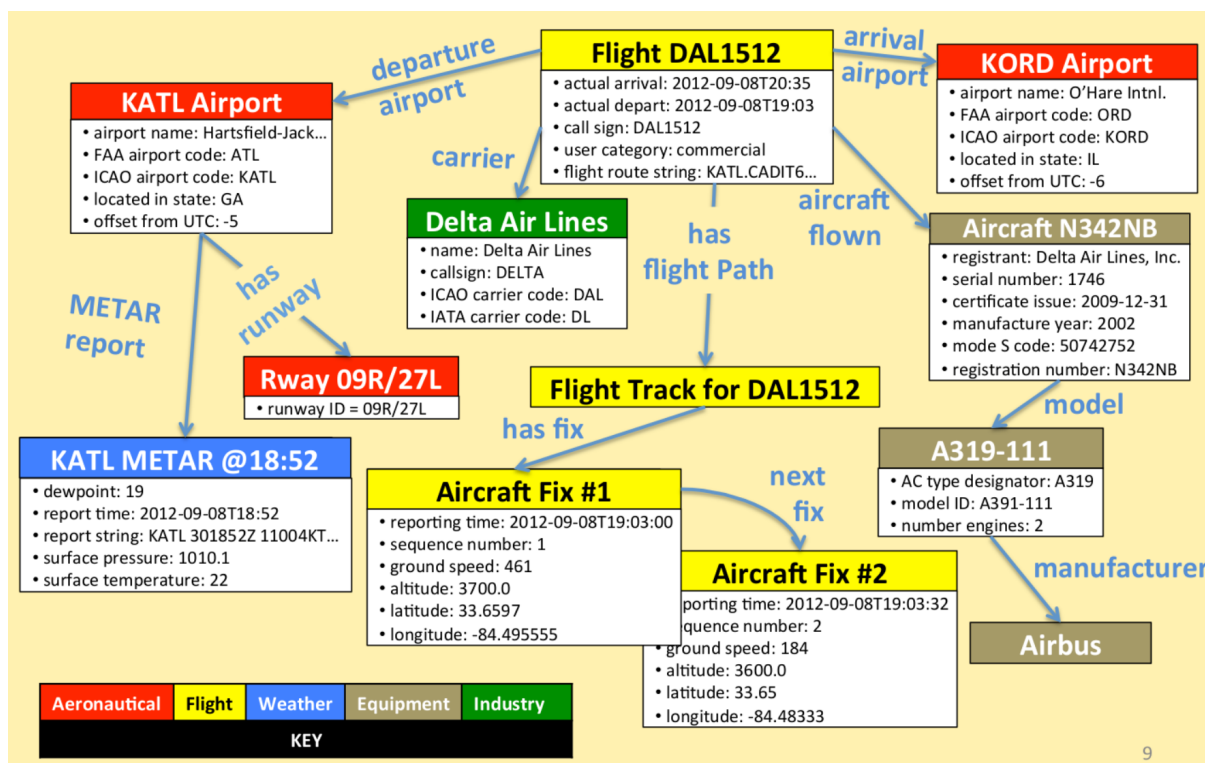


Figura 7.4: Ontologia representando um voo da NASA

De forma semelhante, o AIRM representa um esforço coordenado pela EUROCONTROL para criar um modelo de referência semântico comum para o setor aeronáutico europeu (WILSON; CAHYANI, 2014). O AIRM é derivado de modelos de troca



de informações estabelecidos como padrões para a comunidade aeronáutica global, buscando harmonizar conceitos de modelos específicos como *Aeronautical Information Exchange Model* (AIXM), que padroniza a troca de dados de infraestrutura aeronáutica, *Flight Information Exchange Model* (FIXM), que formaliza a troca de informações de voo, *ICAO Meteorological Information Exchange Model* (IWXXM), que estrutura a troca de informações meteorológicas, e *Airspace Management Exchange Model* (AMXM), que gerencia informações do espaço aéreo.

Esses padrões operam em diferentes níveis de estruturação semântica, desde fundamentos conceituais até implementações técnicas. A Tabela 7.1 apresenta uma síntese desses níveis, ilustrando a relação entre os modelos e suas respectivas funções no ecossistema informacional aeronáutico.

Tabela 7.1: Níveis de Estruturação Semântica no Domínio Aeronáutico

| Nível                      | Exemplo na Aviação      | Analogia                               |
|----------------------------|-------------------------|--|
| Filosofia / Semântica Base | AIRM                    | "A gramática da linguagem aeronáutica" |
| Modelos específicos        | AIXM, FIXM, IWXXM, AMXM | Dicionários técnicos por área          |
| Implementações técnicas    | XML/GML, RDF, APIs      | A forma de escrever documentos reais   |

Ademais, a aplicabilidade de ontologias em sistemas de gerenciamento de dados aeronáuticos é corroborada pela literatura científica de transporte aéreo. Mijović, Babić e Žitnik (2011) apresentam um sistema de apoio à decisão habilitado por ontologia para gerenciamento de emergências em aeroportos, demonstrando como tecnologias semânticas podem processar eventos complexos e fornecer recomendações contextualizadas em situações críticas.

Gringinger et al. (2019) realizaram um estudo comparativo detalhado entre a NASA ATM Ontology e uma ontologia derivada do AIRM. Eles estabeleceram metodologias para alinhamento de ontologias aeronáuticas e identificaram sobreposições conceituais significativas, validando assim a maturidade do domínio para aplicações semânticas.

Assim, as iniciativas internacionais consolidadas fornecem a base conceitual e metodológica necessária para o desenvolvimento da ontologia do **AirData**, adaptada às especificidades do sistema aeronáutico brasileiro.

Por ora, a experiência acumulada por meio das implementações referenciadas permitirá que o projeto **AirData** aproveite padrões semânticos validados internacional-



mente. Isto é, será possível adaptar conceitos para as particularidades regulatórias brasileiras (ANAC, DECEA, SAC), integrar conhecimento contextual específico do transporte aéreo nacional e implementar técnicas comprovadas de integração semântica de dados.

Além dos mais, a convergência das evidências: iniciativas industriais consolidadas, validação científica e necessidades específicas do contexto brasileiro; irá permitir estabelecer que as ontologias constituem o conceito fundamental para viabilizar os objetivos do projeto **AirData**. Elas fornecerão a estrutura semântica necessária para a integração inteligente entre bases de dados analíticos e conhecimento contextual.

## 7.4 Grafos de Conhecimento

A adoção de uma ontologia no projeto **AirData** não se limita à definição conceitual de termos e relações no domínio aeronáutico. Trata-se, sobretudo, da introdução de uma nova sistemática de gestão e governança de dados aeronáuticos brasileiros: dinâmica, conectada e inteligente.

Essa sistemática viabilizará a construção de uma base de dados do tipo Grafo de Conhecimento (*Knowledge Graph*), uma estrutura que permite armazenar, interligar e inferir conhecimento sobre entidades do domínio de forma semanticamente consistente, interoperável e extensível.

No contexto do **AirData**, a ontologia a ser desenvolvida servirá como o esquema semântico que organiza os dados operacionais e normativos. O grafo de conhecimento, por sua vez, será a estrutura computacional que materializa essas relações, conectando fontes heterogêneas (ANAC, DECEA, OpenSky, SAC, REDEMET etc.) e permitindo respostas precisas e explicáveis a questões complexas do transporte aéreo brasileiro.

Os grafos de conhecimento seguem o modelo de representação baseado em triplas semânticas do tipo (sujeito, predicado, objeto), fundamentado em padrões como RDF e OWL. Diferentemente de bancos relacionais tradicionais, os grafos expressam relações explícitas entre entidades (ex.: voos, aeroportos, regulamentos, fenômenos meteorológicos), viabilizando inferências automáticas e consultas semânticas avançadas via SPARQL.

No escopo do **AirData**, os grafos de conhecimento buscarão representar múltiplas dimensões do setor aeronáutico:



- **Conhecimento Estrutural:** relações topológicas entre aeroportos, rotas, espaços aéreos e infraestrutura;
- **Conhecimento Operacional:** dados de voos, atrasos, cancelamentos e eventos em tempo real;
- **Conhecimento Regulatório:** normas ANAC (RBACs), instruções DECEA (ICAs) e regulamentações ICAO vinculadas às entidades operacionais;
- **Conhecimento Contextual:** condições meteorológicas, restrições temporárias e fatores externos que influenciam as operações.

A proposta técnica inicial prevê uma arquitetura híbrida que combine diferentes paradigmas de armazenamento e consulta:

- **Camada RDF/OWL:** assegura conformidade semântica e interoperabilidade com padrões internacionais (AIRM, FIXM, AIXM);
- **Camada Property Graph:** proporciona performance otimizada para análises de conectividade e redes (ex.: com uso de Neo4j);
- **Camada de Sincronização:** mecanismo bidirecional que garante consistência entre representações RDF e Property Graph.

Essa arquitetura considera a escalabilidade necessária ao contexto brasileiro, no qual o sistema deverá lidar com milhares de voos diários e dados de centenas de aeroportos. A camada de sincronização permitirá consultas rápidas (via Neo4j), sem sacrificar a integridade semântica (via RDF), promovendo compatibilidade com sistemas nacionais e internacionais.

A expressividade do modelo pode ser ilustrada pela consulta abaixo, que identifica conexões com escala entre Frankfurt (FRA) e Seattle (SEA):

```
PREFIX prop: <http://kelvinlawrence.net/air-routes/vocab/prop#>
PREFIX airport: <http://kelvinlawrence.net/air-routes/data/airport#>
SELECT DISTINCT ?via WHERE {
  ?route1 prop:from airport:FRA .
  ?route1 prop:to ?via .
  ?route2 prop:from ?via .
  ?route2 prop:to airport:SEA .
}
```



A Figura 7.5 ilustra visualmente o resultado da consulta. Os aeroportos (nós azuis) estão conectados por múltiplos tipos de relacionamento (rotas, localização, operações), evidenciando como conceitos geográficos e operacionais podem ser integrados semanticamente.

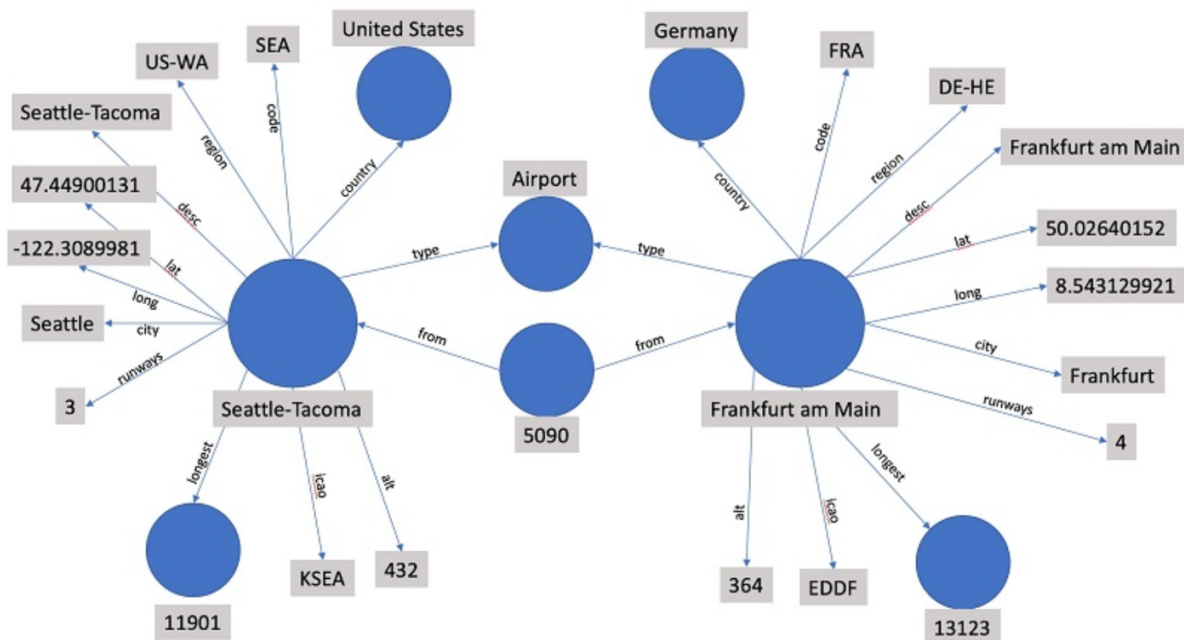


Figura 7.5: Exemplo de estrutura de um Grafo de Conhecimento aplicado à aviação

Para o contexto nacional, consultas específicas poderão, por exemplo, identificar efeitos em cascata de atrasos entre voos, com base em relações semânticas explícitas:

```
PREFIX \airdata: <http://\airdata.org/ontologia#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?voo_afetado ?atraso_cascata ?justificativa ?norma WHERE {
    ?voo_origem \airdata:atrasoPor ?causa_inicial ;
                \airdata:origem \airdata:SBGR ;
                \airdata:atrasoMinutos ?atraso_original .
    ?voo_afetado \airdata:dependeDe ?voo_origem ;
                \airdata:atrasoCascata ?atraso_cascata ;
                rdfs:label ?numero_voo .
    ?causa_inicial \airdata:referenciaNorma ?norma ;
                rdfs:comment ?justificativa .
    FILTER(?atraso_original > 15)
}
```



Além do armazenamento e consulta, o grafo de conhecimento permitirá inferências automáticas baseadas em regras semânticas, como:

- Inferir restrições operacionais quando a visibilidade for inferior a 600 m (com base em ICAs);
- Inferir direitos do passageiro em voos com atrasos superiores a 4 h (segundo regulamentação ANAC);
- Detectar causas estruturais de atrasos recorrentes em determinadas rotas;
- Correlacionar eventos meteorológicos com seus impactos operacionais.

A adoção de grafos de conhecimento como base para integração semântica de dados tem sido progressivamente consolidada na literatura científica especializada. Ghafari, Tan e Rezgui (2023) demonstram a aplicabilidade desse paradigma para integração e modelagem de dados complexos em engenharia civil, evidenciando sua capacidade de representação semântica em domínios de alta complexidade técnica e regulatória.

No contexto de metodologias de desenvolvimento, Berardi e Auceli (2024) enfatizam a relevância do envolvimento sistemático de especialistas de domínio nos processos de modelagem conceitual, enquanto Rezgui, Ramonell e Tan (2022) propõem uma metodologia axiomática para o projeto estruturado de grafos de conhecimento em ambientes de Indústria 4.0.

Especificamente no domínio aeronáutico, Cahyani e Wilson (2023) apresentam uma arquitetura de grafo de conhecimento voltada à comunidade internacional de gerenciamento de tráfego aéreo, demonstrando a viabilidade técnica e conceitual dessa abordagem para o transporte aéreo.

De particular relevância para o projeto **AirData**, Sharma e Meena (2023) propõem um modelo fundamentado em ontologias para alimentação automatizada de grafos de conhecimento com dados do setor de transporte aéreo, mediado por agentes conversacionais inteligentes. Essa abordagem viabiliza a integração sistemática de conhecimento contextual e operacional a partir de entradas não estruturadas, evidenciando o potencial dos grafos para capturar e operacionalizar conhecimento tácito especializado no domínio aeronáutico.

As evidências apresentadas corroboram tanto a maturidade técnica quanto a aderência conceitual dos grafos de conhecimento como solução estruturante para os objetivos do projeto **AirData**. Sua adoção possibilitará não apenas a interoperabilidade



semântica entre sistemas e fontes de dados heterogêneas, mas também a construção de uma arquitetura cognitiva capaz de inferir, justificar e adaptar-se dinamicamente às demandas regulatórias, operacionais e contextuais específicas do sistema de transporte aéreo brasileiro.

## 7.5 Base de Conhecimento Semântica

Para que o **AirData** atinja seus objetivos, será necessário, de forma complementar ao desenvolvimento das ontologias e dos grafos de conhecimento, planejar e implementar uma base de conhecimento semântica. Essa base terá como finalidade organizar, de maneira estruturada, o conhecimento técnico, normativo e operacional do setor de transporte aéreo brasileiro, considerando também referências internacionais relevantes, como as da ICAO e da IATA.

A base será projetada para alimentar os *Large Language Models* (LLM) e fornecer respostas contextualizadas a questionamentos sobre causas, normas e implicações de eventos operacionais. No entanto, esse processo envolve vários desafios, dado o alto volume, a diversidade e a constante atualização dos documentos regulatórios e técnicos do setor aeronáutico - o que exigirá estratégias eficientes de governança e manutenção para evitar altos custos de reprocessamento e readequação de modelos.

Para enfrentar esses desafios, propõe-se a adoção de uma metodologia estruturada que transforme documentos não estruturados em uma base de conhecimento semanticamente enriquecida e consultável. A Tabela 7.2 resume as etapas iniciais previstas para essa construção.



Tabela 7.2: Etapas Iniciais da Construção da Base de Conhecimento Semântica

| Etapa                                    | Descrição   | Resultado Esperado  |
|--|---|---|
| Coleta e Organização do Corpus           | Reunião sistemática de documentos regulatórios, técnicos e operacionais de múltiplas fontes (ANAC, DECEA, ICAO etc.), organizados em diretórios por origem e tipo | Repositório documental estruturado                                |
| Processamento e Enriquecimento Semântico | Conversão para texto, extração de metadados, e anotação de entidades relevantes com base em técnicas de NLP   | Base textual enriquecida semanticamente                           |
| Geração de Vetores (Embeddings)          | Transformação dos trechos textuais em vetores numéricos que representam seus significados semânticos  | Representação vetorial do conhecimento                            |
| Armazenamento e Consulta em Vector Store | Inserção dos embeddings em base especializada para buscas semânticas de alta performance  | Sistema capaz de responder a perguntas por similaridade semântica |
| Mapeamento para Ontologia                | Vinculação das entidades textuais às classes e propriedades da ontologia do projeto <b>AirData</b>  | Integração semântica entre textos e grafo de conhecimento         |

A seguir, detalham-se as etapas planejadas para a construção da base.

### *Etapa 1: Coleta e Organização do Corpus*

A etapa inicial consistirá na reunião sistemática do corpus documental, que deverá incluir:

- Regulamentos emitidos pela ANAC;
- Instruções e normas do DECEA e do Comando da Aeronáutica;
- Manuais técnicos, circulares e documentos da ICAO;
- Legislação nacional aplicável, como o RBAC;
- Normas de associações internacionais como IATA e EUROCONTROL.

Os documentos deverão ser organizados por origem, tipo e data, compondo um repositório centralizado e padronizado.



## Etapa 2: Processamento e Enriquecimento Semântico

Nesta etapa, os documentos serão convertidos e enriquecidos com metadados estruturados por meio do seguinte fluxo:

1. **Conversão para texto:** Arquivos PDF, DOC e outros formatos serão transformados em texto limpo, com uso de *Optical Character Recognition* (OCR) quando necessário;
2. **Metadados estruturados:** Serão adicionados atributos como título, órgão emissor, tipo de documento, data de publicação e versão;
3. **Anotação de entidades:** Técnicas de *Natural Language Processing* (NLP) serão aplicadas para identificar entidades relevantes, como *aeroporto*, *cancelamento* ou *visibilidade reduzida*.

Como resultado, espera-se obter uma base documental com textos indexados e enriquecidos semanticamente.

## Etapa 3: Geração de Vetores (Embeddings)

Nesta fase, trechos relevantes (parágrafos, artigos ou instruções) serão transformados em vetores numéricos (*embeddings*), representando o conteúdo semântico em um espaço vetorial. Isso permitiria buscas por similaridade de significado, indo além da simples correspondência por palavras-chave. Ferramentas como FAISS, Weaviate ou modelos de embedding da OpenAI ou de código aberto serão avaliadas para essa função.

## Etapa 4: Armazenamento em Vector Store e Consulta Semântica

Os vetores serão armazenados em uma base especializada do tipo *vector store*, projetada para buscas de alta performance por similaridade semântica. Com isso, será possível, por exemplo, localizar trechos normativos que tratem de “*regras para cancelamento por nevoeiro*”, mesmo que os termos exatos não estejam presentes no texto original.



## Etapa 5: Mapeamento Conceitual para a Ontologia

A etapa final compreenderia o mapeamento das entidades extraídas da base de conhecimento para as classes e propriedades definidas na ontologia do **AirData**. Exemplos de mapeamentos planejados:

- Um trecho sobre “nevoeiro” e “visibilidade” poderia ser vinculado à entidade :CausaAtraso na ontologia;
- A norma “ICA 100-12” seria representada como instância da classe :Norma.

Esse mapeamento possibilitará que o sistema responda perguntas combinando os dados do grafo analítico com o conhecimento normativo extraído da base textual, promovendo explicabilidade e contextualização nas respostas.

### 7.6 Convergência e Geração de Respostas

As seções anteriores descreveram as duas trilhas técnicas propostas para o projeto **AirData**: a criação de um Grafo de Conhecimento, voltado à organização de dados operacionais estruturados, e o desenvolvimento de uma Base de Conhecimento Semântica, dedicada ao conhecimento normativo e contextual. Porém, para que o **AirData** alcance plenamente seus objetivos, será necessário implementar um processo de convergência entre essas duas camadas.

Portanto, o objetivo desta nova etapa é capacitar o sistema a responder a perguntas complexas, tanto aquelas que demandam dados factuais quanto contexto regulatório, de forma unificada, confiável e explicável.

A questão principal é: *como permitir que o sistema, a partir de uma única pergunta, acesse duas fontes distintas de informação e produza uma resposta coerente?* A solução proposta baseia-se na adoção de uma arquitetura de Inteligência Artificial conhecida como *Retrieval-Augmented Generation (RAG)*. Inspirado no comportamento de um especialista humano, o RAG opera combinando duas capacidades:

1. **Recuperação de fatos estruturados:** Assim como um especialista recorre à memória para lembrar dados operacionais, o sistema consulta o Grafo de Conhecimento.



2. **Consulta a fontes normativas:** O especialista também sabe onde buscar a fundamentação normativa ou técnica para interpretar um evento. O sistema faz isso consultando a Base de Conhecimento Semântica.

O RAG automatiza esse processo: ele primeiro recupera dados confiáveis e contextos relevantes e, a seguir, utiliza um Modelo LLM para gerar uma resposta explicada e contextualizada. Seja pergunta ilustrativa do presente relatório: *Por que o voo LATAM 5411 atrasou?* A geração da resposta segue quatro etapas principais, ilustradas na Figura 7.6:

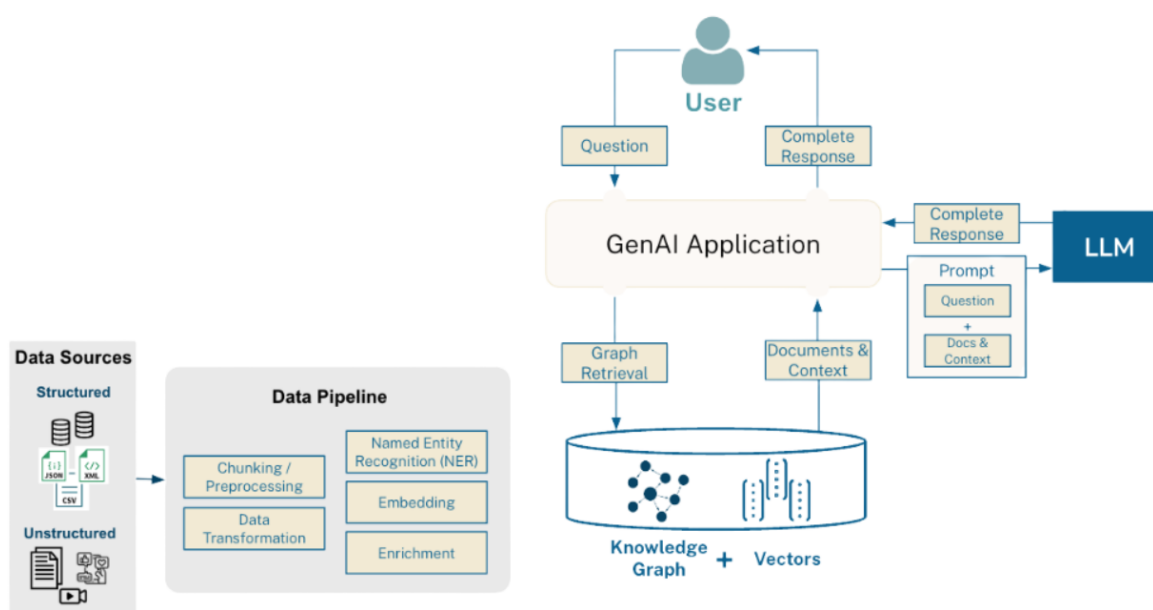


Figura 7.6: Fluxo ilustrativo da arquitetura RAG no contexto **AirData**. Fonte: Matos (2024)

No contexto da Figura 7.6, tem-se:

1. **Recuperação de Fatos Estruturados:** A aplicação consulta o Grafo de Conhecimento para identificar dados objetivos do evento:
  - **Resultado:** “O voo LATAM 5411 (SBGR–SBPA) teve um atraso de 45 minutos.” O sistema obtém informações como quem, o quê, quando e onde.
2. **Recuperação de Conhecimento Contextual:** Simultaneamente, realiza-se uma busca semântica na Base de Conhecimento Vetorial, retornando trechos que mencionam “condições meteorológicas”, “nevoeiro em Guarulhos” e a norma “ICA 100-12”.



3. **Montagem do Prompt + Geração Aumentada:** O sistema envia os resultados ao LLM com a instrução: “Com base nestes fatos e neste contexto normativo, elabore uma resposta à pergunta.”
4. **Síntese da Resposta:** O modelo gera uma resposta explicável, como:
  - **Resposta:** “O voo LATAM 5411 (SBGR–SBPA) teve atraso de 45 minutos devido a nevoeiro em Guarulhos. Segundo a ICA 100-12, visibilidade reduzida exige maior espaçamento entre aeronaves, reduzindo a capacidade do aeroporto e gerando atrasos operacionais.”

Essa convergência técnica traz benefícios decisivos:

- **Fundamentação:** As respostas são ancoradas em dados e documentos auditáveis, e não em conhecimento genérico da internet.
- **Explicabilidade:** O sistema pode apontar os trechos normativos e dados operacionais que justificam suas conclusões.
- **Confiabilidade:** Minimiza o risco de *alucinações* - respostas imprecisas, comuns em IAs não conectadas a fontes verificáveis.

Em resumo, a adoção da arquitetura RAG permitirá que o **AirData** evolua para uma plataforma cognitiva, que integra raciocínio automatizado e conhecimento especializado, oferecendo respostas confiáveis, contextualizadas e transparentes aos gestores do setor aéreo brasileiro.

## 7.7 Benefícios Técnicos e Ganhos Operacionais

A seguir, apresentam-se os principais benefícios operacionais decorrentes das propostas técnicas do projeto **AirData**, com impactos diretos para a ANAC, DECEA e SAC:

1. **Ontologia Comum:** Estabelece um vocabulário padronizado e formal para variáveis e conceitos operacionais, eliminando ambiguidades na interpretação de termos como “cancelamento”, “aeronave em solo” ou “visibilidade reduzida”. Essa padronização semântica irá permitir a integração confiável entre dados de diferentes fontes e melhora a consistência de indicadores estratégicos utilizados por ANAC, DECEA e SAC.



2. **Grafo de Conhecimento:** Permite análises de causa e efeito complexas, como identificar o impacto em cascata de atrasos em um aeroporto para toda a malha aérea nacional, apoiando o planejamento tático do DECEA.
3. **Consultas SPARQL:** Viabiliza a criação de painéis de controle dinâmicos e personalizados que respondem a perguntas de negócio complexas, sem exigir o desenvolvimento de novas rotinas de ETL a cada nova análise.
4. **Raciocínio Automático:** Automatiza a verificação de conformidade de eventos operacionais com as normas vigentes, acelerando processos de fiscalização e auditoria conduzidos pela ANAC.
5. **Arquitetura RAG:** Gera respostas explicáveis e fundamentadas, combinando dados operacionais com normas em tempo real. Fortalece a tomada de decisão baseada em evidências, aumenta a confiabilidade institucional e reduz o tempo de resposta frente a eventos críticos.

## 7.8 Considerações Finais

As propostas apresentadas neste capítulo estabelecem uma metodologia para que o **AirData** evolua de um conceito para uma solução prática, escalável e aderente às necessidades da CNJ, ANAC, DECEA e SAC. A integração entre bases de dados analíticos e conhecimento semântico, sustentada por ontologias e grafos de conhecimento, fornece a infraestrutura necessária para que análises complexas sejam conduzidas de forma confiável, explicável e em tempo hábil.

A adoção da arquitetura RAG amplia o potencial da plataforma, permitindo que dados operacionais e conhecimento normativo sejam consultados de maneira integrada, gerando respostas contextualizadas e fundamentadas em evidências verificáveis. Isso representa não apenas um ganho tecnológico, mas também um avanço institucional na forma como informações críticas são tratadas e utilizadas para tomada de decisão.

Com a aplicação coordenada dessas abordagens, o **AirData** se será um modelo de referência para a interoperabilidade de dados no setor de transporte aéreo, capaz de alinhar inovação tecnológica, conformidade regulatória e eficiência operacional. As próximas ações do projeto serão focadas na implementação gradual dessas camadas, assegurando que os resultados sejam validados por especialistas e que o sistema mantenha flexibilidade para incorporar futuras evoluções do domínio aeronáutico.



## 8 Considerações Finais

O projeto **AirData** representa uma proposta estruturante para a modernização da gestão de dados no setor de transporte aéreo brasileiro, combinando metodologias consolidadas de integração semântica com tecnologias emergentes de inteligência artificial. Ao longo deste relatório, foram apresentadas soluções técnicas capazes de superar limitações históricas na interoperabilidade entre bases de dados e no acesso a conhecimento contextualizado, estabelecendo um caminho concreto para a criação de uma plataforma unificada, escalável e aderente às normas do setor.

Com base nas propostas técnicas delineadas, o **AirData** encontra-se conceitualmente estruturado e tecnicamente fundamentado para iniciar sua fase de implementação. A seguir, apresentam-se os próximos passos recomendados para a continuidade do desenvolvimento:

1. **Modelagem Inicial da Ontologia AirData:** Formalização do vocabulário comum por meio de um modelo OWL inicial, priorizando conceitos operacionais críticos (voo, atraso, aeroporto, norma).
2. **Desenho e Implementação do Grafo de Conhecimento:** Construção da infraestrutura de dados baseada em RDF e Property Graph, com ingestão piloto de dados operacionais (ex.: voos históricos, condições meteorológicas).
3. **Coleta e Curadoria do Corpus Normativo:** Reunião sistemática de documentos regulatórios e técnicos relevantes, incluindo normas ANAC, ICAs do DECEA e documentos da ICAO, organizados em repositório temático.
4. **Pipeline de Vetorização e Indexação Semântica:** Desenvolvimento de mecanismos de extração de conhecimento, geração de embeddings e armazenamento em *vector store*, com integração à ontologia.
5. **Integração Arquitetural com Framework RAG:** Prototipação da arquitetura de geração aumentada, conectando o grafo analítico, a base semântica vetorial e modelos LLM ajustados ao domínio aeronáutico.
6. **Validação com Casos de Uso Prioritários:** Definição de perguntas representativas e execução de testes com usuários especialistas (ANAC, CNJ, DECEA, SAC), avaliando acurácia, confiabilidade e explicabilidade das respostas.



- 7. Planejamento da Governança e Escalabilidade:** Estabelecimento de mecanismos para atualização contínua da base de conhecimento, versionamento normativo e expansão modular do sistema **AirData**.

A execução coordenada desses passos permitirá não apenas a materialização da arquitetura proposta, mas também a criação de um ambiente propício à inovação contínua, no qual novas fontes de dados e métodos analíticos poderão ser incorporados sem comprometer a integridade semântica e a coerência operacional do sistema.

Por fim, a continuidade bem-sucedida do **AirData** dependerá da articulação entre técnicos do ITA e os técnicos institucionais da ANAC, DECEA e SAC, garantindo não apenas a implementação tecnológica, mas também a apropriação e incorporação do sistema como ferramenta estratégica no ecossistema da aviação civil brasileira. Ao assegurar alinhamento institucional, robustez técnica e governança efetiva, o **AirData** tem potencial para se tornar uma referência nacional em integração de dados e inteligência aplicada ao transporte aéreo.



## Bibliografia

ANAC. **Resolução nº 191, de 16 de junho de 2011**. 2011. <<https://www.anac.gov.br/assuntos/legislacao/legislacao-1/resolucoes/resolucoes-2011/resolucao-no-191-de-16-06-2011>>. Acesso em: 25 ago. 2025.

ANAC. **Resolução nº 440, de 9 de agosto de 2017**. 2017. <<http://www.anac.gov.br/assuntos/legislacao/legislacao-1/resolucoes/2017/resolucao-no-440-09-08-2017>>. Acesso em: 07 jul. 2025.

ANAC. **Manual do Sistema de Envio dos Dados Estatísticos de Voos (DataVoo)**. 2019. <<https://sas.anac.gov.br/sas/estatistica/lib/manual.pdf>>. Acesso em: 25 ago. 2025.

ANAC. **Portaria nº 3.506/SAS, de 11 de novembro de 2019**. 2019. <<https://www.anac.gov.br/assuntos/legislacao/legislacao-1/portarias/2019/portaria-no-3506-sas-11-11-2019>>. Acesso em: 25 ago. 2025.

ANAC. **Portaria nº 3.507/SAS, de 11 de novembro de 2019**. 2019. <[https://www.anac.gov.br/assuntos/legislacao/legislacao-1/portarias/2019/portaria-no-3507-sas-11-11-2019/@@display-file/arquivo\\_norma/PA2019-3507%20-%20Compilado%20at%C3%A9%20PA2019-3902.pdf](https://www.anac.gov.br/assuntos/legislacao/legislacao-1/portarias/2019/portaria-no-3507-sas-11-11-2019/@@display-file/arquivo_norma/PA2019-3507%20-%20Compilado%20at%C3%A9%20PA2019-3902.pdf)>. Acesso em: 25 ago. 2025.

ANAC. **Portaria nº 2.177/SAS, de 26 de agosto de 2020**. 2020. <<https://www.anac.gov.br/assuntos/legislacao/legislacao-1/portarias/2020/portaria-no-2177-sas-26-08-2020>>. Estabelece os procedimentos para o registro prévio de serviços de transporte aéreo público.

ANAC. **Descrição das variáveis (microdados)**. 2023. <<https://www.gov.br/anac/pt-br/assuntos/dados-e-estatisticas/mercado-do-transporte-aereo/demanda-e-oferta/descricao-das-variaveis-microdados>>. Acesso em: 25 ago. 2025.

ATC, S. **Total Air Traffic Information Control - TATIC**. 2020. Disponível em: <<https://www.saipher.com.br/tatic-twr/>>. Acesso em: 24 jul. 2025.

BERARDI, A.; AUCELI, J. Construção de grafos de conhecimento baseados em dados não-estruturados com envolvimento de especialistas de domínio. In: SOCIEDADE BRASILEIRA DE COMPUTAÇÃO (SBC). **Anais do XIX Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2024)**. [S.l.], 2024.

BERNERS-LEE, T. **Open Data – Three Rules for Government Data**. 2009. <<https://www.w3.org/DesignIssues/GovData.html>>. Accessed: 2025-08-10.

BRASIL. **Lei nº 7.183, de 5 de abril de 1984: Lei do Aeronauta**. 1984. <[https://www.planalto.gov.br/ccivil\\_03/leis/l7183.htm](https://www.planalto.gov.br/ccivil_03/leis/l7183.htm)>. Dispõe sobre o exercício da profissão de aeronauta e dá outras providências.



BRASIL. **Lei nº 7.565, de 19 de dezembro de 1986: Código Brasileiro de Aeronáutica.** 1986. <[https://www.planalto.gov.br/ccivil\\_03/leis/l7565.htm](https://www.planalto.gov.br/ccivil_03/leis/l7565.htm)>. Dispõe sobre o Código Brasileiro de Aeronáutica e dá outras providências.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011: Lei de Acesso à Informação.** 2011. <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm)>. Regula o acesso a informações previsto na Constituição Federal e dá outras providências.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018: Lei Geral de Proteção de Dados Pessoais (LGPD).** 2018. <[https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm)>. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet).

CAHYANI, D.; WILSON, S. Building a knowledge graph for the air traffic management community. In: ICRAT. **Proceedings of the International Conference on Research in Air Transportation (ICRAT).** [S.l.], 2023.

CWIEK, P. **pmetar: Processing METAR Weather Reports.** [S.l.], 2025. Acessado em: 23 jul. 2025. Disponível em: <<https://github.com/prcwiek/pmetar>>.

DECEA. **Publicação nº CIRTRAF 100-10, de 1º de julho de 2006.** 2006. <<https://publicacoes.decea.gov.br/?i=publicacao&id=2738>>. Acesso em: 24 jul. 2025.

DECEA. **ICA 100-37: Serviços de Tráfego Aéreo.** 2024. <<https://publicacoes.decea.mil.br/publicacao/ica-100-37>>. Em vigor desde 28 nov. 2024. Acesso em: ago. 2025.

FR24. **FlightRadar24 - Airports by country.** 2025. <<https://www.flightradar24.com/data/airports>>. Acesso em: jul. 2025.

GHAFFARI, R.; TAN, H.; REZGUI, Y. Knowledge graph-based data integration system for digital twins of built assets. **Advanced Engineering Informatics**, Elsevier, v. 57, p. 101188, 2023.

Google LLC. **Google Cloud – Cloud Computing Services & APIs.** 2025. <<https://cloud.google.com/?hl=pt-br>>. Acessado em: 23 jul. 2025.

GRINGINGER, E. et al. Ontology alignment for air traffic management: Comparing nasa atm ontology and airm. In: **Proceedings of the Integrated Communications, Navigation and Surveillance Conference (ICNS).** [S.l.]: IEEE, 2019. p. 1–9.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, v. 5, n. 2, p. 199–220, 1993.

INMET. **INMET – Sobre o Instituto Nacional de Meteorologia.** 2025. <<https://portal.inmet.gov.br/sobre>>. Acessado em: 23 jul. 2025.

KELLER, J. M.; JOBE, K.; MULFINGER, D. An ontology for the management of national airspace system performance. In: IEEE. **Proceedings of the Integrated Communications, Navigation and Surveillance (ICNS) Conference.** [S.l.], 2017. p. 1–12.



KEMP, M. U. et al. **RNCEP: Obtain, Organize, and Visualize NCEP Weather Data**. R Foundation for Statistical Computing, 2025. Acessado em: 23 jul. 2025. Disponível em: <<https://cran.r-project.org/web/packages/RNCEP/RNCEP.pdf>>.

MATOS, D. **IA Generativa – GraphRAG x RAG**. 2024. Acesso em: 10 ago. 2025. Disponível em: <<https://www.cienciaedados.com/ia-generativa-graphrag-x-rag/>>.

MIJOVIĆ, S.; BABIĆ, D.; ŽITNIK, S. Ontology-based decision support system for emergency management in airports. In: **Proceedings of the 14th International Conference on Transport Science (ICTS)**. [S.l.]: University of Ljubljana, Faculty of Maritime Studies and Transport, 2011. p. 1–10.

National Centers for Environmental Prediction (NCEP). **NCEP – National Centers for Environmental Prediction**. 2025. <<https://www.weather.gov/ncep/>>. Acessado em: 23 jul. 2025.

REZGUI, Y.; RAMONELL, R.; TAN, H. Rdf knowledge graphs designed with axiomatic methodology to enhance industry 4.0. **Journal of Industrial Information Integration**, v. 27, p. 100293, 2022.

SHARMA, V.; MEENA, H. Ontology based models of chatbots for populating knowledge graphs. In: IEEE. **2023 International Conference on AI in Transportation Systems (ICAITS)**. [S.l.], 2023.

WILSON, S.; CAHYANI, D. The atm information reference model (airm): Status and plans. In: IEEE/AIAA. **Proceedings of the 31st Digital Avionics Systems Conference (DASC)**. [S.l.], 2014. p. 1C2–1–1C2–9.

