

UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS GRADUAÇÃO EM FÍSICA AMBIENTAL

**ESTRATÉGIA COMPUTACIONAL PARA
REPRODUTIBILIDADE E
ARMAZENAMENTO DE DADOS
AMBIENTAIS**

Igor Antonio Kuhnen

Orientador: **Prof. Dr. José de Souza Nogueira**
Coorientador: **Prof. Dr. Rafael Da Silva Palácios**

Cuiabá - MT
Agosto/2020

UNIVERSIDADE FEDERAL DE MATO GROSSO
INSTITUTO DE FÍSICA
PROGRAMA DE PÓS GRADUAÇÃO EM FÍSICA AMBIENTAL

ESTRATÉGIA COMPUTACIONAL PARA
REPRODUTIBILIDADE E
ARMAZENAMENTO DE DADOS
AMBIENTAIS

Igor Antonio Kuhnen

Tese apresentada ao Programa de Pós Graduação
em Física Ambiental da Universidade Federal de
Mato Grosso, como parte dos requisitos para ob-
tenção do título de Doutor em Física Ambiental.

Orientador: **Prof. Dr. José de Souza Nogueira**

Coorientador: **Prof. Dr. Rafael Da Silva Palácios**

Cuiabá, MT

Agosto/2020

Dados Internacionais de Catalogação na Fonte.

K96e Kuhnen, Igor Antonio.
ESTRATÉGIA COMPUTACIONAL PARA REPRODUTIBILIDADE E
ARMAZENAMENTO DE DADOS AMBIENTAIS / Igor Antonio Kuhnen. -- 2020
58 f. : il. color. ; 30 cm.

Orientador: José de Souza Nogueira.

Co-orientador: Rafael da Silva Palácios.

Tese (doutorado) - Universidade Federal de Mato Grosso, Instituto de Física,
Programa de Pós-Graduação em Física Ambiental, Cuiabá, 2020.

Inclui bibliografia.

1. Dados ambientais. 2. Banco de dados. 3. Reprodutibilidade. 4. Laravel. I.
Título.

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

Permitida a reprodução parcial ou total, desde que citada a fonte.



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE MATO GROSSO
PRÓ-REITORIA DE ENSINO DE PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA AMBIENTAL

FOLHA DE APROVAÇÃO

TÍTULO: ESTRATÉGIA COMPUTACIONAL PARA REPRODUTIBILIDADE E ARMAZENAMENTO DE DADOS AMBIENTAIS

AUTOR (A): DOUTORANDO IGOR ANTONIO KUHNEN

Tese defendida e aprovada em **11 de agosto de 2020**.

COMPOSIÇÃO DA BANCA EXAMINADORA

1. Prof. Dr. José de Souza Nogueira - Orientador – Instituto de Física/UFMT
2. Prof. Dr. Rafael da Silva Palácios - Coorientador - Instituto de Geociências/Universidade Federal do Pará/UFPA
3. Prof. Dr. Leone Francisco Amorim Curado – Examinador Interno - – Instituto de Física/UFMT
4. Pro Dr. Gerson Kazuyoshi Kida – Examinador Externo – Instituto Federal de Mato Grosso/UFMT
5. Prof. Dr. Denes Martins de Moraes – Examinador Externo - Centro Universitário de Várzea Grande – UNIVAG

Cuiabá-MT, 11 de agosto de 2020



Documento assinado eletronicamente por **SERGIO ROBERTO DE PAULO, Docente da Universidade Federal de Mato Grosso**, em 11/08/2020, às 18:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **JOSE DE SOUZA NOGUEIRA, Docente da Universidade Federal de Mato Grosso**, em 11/08/2020, às 18:14, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **LEONE FRANCISCO AMORIM CURADO, Docente da Universidade Federal de Mato Grosso**, em 11/08/2020, às 18:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Gerson Kazuyoshi Kida, Usuário Externo**, em 11/08/2020, às 22:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rafael da Silva Palácios, Usuário Externo**, em 12/08/2020, às 09:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Denes Martins de Moraes, Usuário Externo**, em 12/08/2020, às 19:13, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufmt.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2732470** e o código CRC **AFE2BD46**.

DEDICATÓRIA

*À Karyna, Théo e Íris, esposa e filhos,
pelo carinho, atenção e pelo sacrifício
das horas mais difíceis.*

Agradecimentos

Primeiramente a Deus, por permitir fazer parte desse programa, no qual tenho a oportunidade de avançar em soluções relevantes para a comunidade científica e a sociedade.

A minha família e meus amigos por todo o suporte e incentivo em todos os momentos. O que proporcionou forças para enfrentar os desafios diários.

Ao meu orientador Dr. José de Souza Nogueira pela acolhida, conselhos e apoio incondicional em toda essa jornada.

Ao coorientador Rafael da Silva Palácios, pelas dicas, instruções, conversas, etc.

Ao amigo Lucas Rothmund pelas discussões, argumentações levantadas no trabalho.

Gostaria de agradecer a todas as pessoas do PGFA que fizeram parte deste trabalho.

À Capes - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - pela concessão de bolsa de doutorado.

SUMÁRIO

LISTA DE FIGURAS	I
LISTA DE SÍMBOLOS	III
RESUMO	IV
ABSTRACT	V
1 INTRODUÇÃO	1
1.1 PROBLEMÁTICA	1
1.2 JUSTIFICATIVA	2
1.3 OBJETIVO GERAL	2
1.4 OBJETIVOS ESPECÍFICOS	3
2 FUNDAMENTAÇÃO TEÓRICA	4
2.1 DADOS AMBIENTAIS	4
2.2 GESTÃO DOS DADOS	6
2.2.1 Gestão de Dados Científicos	7
2.2.2 Ciclo de Vida dos Dados	10
2.2.2.1 O Ciclo de Vida dos Dados Tradicional	10
2.2.2.2 O Novo Ciclo de Vida dos Dados	11
2.3 BANCO DE DADOS	12
2.3.1 Replicação de Dados	13
2.4 LINGUAGEM DE MODELAGEM PADRONIZADA (<i>Unified Modeling Language-UML</i>)	14
2.4.1 Diagrama de Entidade Relacionamento (DER)	15

2.4.2	Diagrama de Atividades	15
2.4.3	Diagrama de Fluxo de Dados (DFD)	16
3	MATERIAL E MÉTODOS	18
3.1	LEVANTAMENTO AMBIENTAL SOBRE O PANTANAL MATO- GROSSENSE	18
3.1.1	Local de Origem dos Estudos	18
3.1.2	Aetalômetro	19
3.1.3	Nefelômetro	20
3.1.4	Descrição do Processo de Coleta	21
3.2	DIAGRAMAS DO SISTEMA	22
3.3	DESENVOLVIMENTO DA FERRAMENTA WEB	24
3.3.1	Requisitos	24
3.3.1.1	Requisitos Computacionais	24
3.3.1.2	Requisitos de Conteúdo	25
3.3.1.3	Requisitos de Gerenciamento	25
3.3.1.4	Requisitos de Uso	26
3.3.2	Ferramentas de Programação	26
3.3.2.1	Editor de Código Fonte	26
3.3.2.2	Linguagem de Programação	26
3.3.2.3	<i>Framework</i>	27
3.3.2.4	Sistema Gerenciador de Banco de Dados	28
3.4	REPLICABILIDADE DOS DADOS	28
4	RESULTADOS	29
4.1	DESENVOLVIMENTO DO SISTEMA	29
4.2	APRESENTAÇÃO DO AMBIENTE	34
4.3	FUNCIONALIDADES DO AMBIENTE	37
5	CONCLUSÃO	40
5.1	TRABALHOS FUTUROS	41
	REFERÊNCIAS	42

LISTA DE FIGURAS

1	Fluxo de atividades no monitoramento ambiental.	5
2	Dados, informação e conhecimento	6
3	Representação de entropia dos dados científicos.	8
4	Esquema apresentando a cauda longa dos dados científicos.	9
5	Representação do antigo ciclo de vida dos dados.	10
6	Representação do novo ciclo de vida dos dados.	11
7	Representação do ciclo de vida dos dados científicos.	12
8	Replicação de dados.	14
9	Simbologia das conexões do DER.	15
10	Simbologia das conexões do Diagrama de Atividades.	16
11	Simbologia das conexões do DFD.	17
12	Região do pantanal matogrossense onde encontra-se a BAPP-UFMT, local onde encontra-se os equipamentos da retirada dos dados para desenvolvimento do sistema.	19
13	Aetalômetro utilizado na BAPP.	20
14	Princípio de funcionamento do aetalômetro.	20
15	Nefelômetro utilizado na BAPP.	21
16	Representação do Processo Antigo do processamento dos dados.	21
17	Representação do Processo Automatizado do processamento dos dados.	23
18	Diagrama Entidade Relacionamento onde apresenta a estrutura das tabelas criadas no banco de dados.	24

19	Linhas de códigos das rotas do sistema.	30
20	Comando para criação da Migration.	30
21	Lista das <i>migrations</i> criadas.	31
22	Parte do código de uma <i>migration</i>	31
23	Comando para criação das <i>Models</i>	32
24	Parte do código de uma <i>Model</i>	32
25	Parte do código de uma <i>View</i>	33
26	Tela de apresentação do sistema.	34
27	Gráfico de Concentração de <i>Black Carbon</i> apresentado pelo Grafana.	35
28	Gráfico de Coeficiente de Espalhamento apresentado pelo Grafana.	36
29	Gráfico de Coeficiente de Espalhamento apresentado pelo Grafana.	36
30	Tela de apresentação e descrição dos equipamentos na BAPP.	37
31	Tela de login do sistema para acesso aos dados do sistema.	38
32	Tela de escolha dos dados a serem visualizados para cada equipamento.	38
33	Tela de exibição de todos os dados do equipamento selecionado no banco de dados.	38
34	Tela para inserção do novo dados, caso seja necessária alguma manutenção.	39
35	Tela que apresenta o botão para importação dos dados que estavam no equipamento para o banco de dados. A exportação é o <i>download</i> de todos os dados que esteja no banco de dados para arquivo de saída.	39

LISTA DE SÍMBOLOS

ATN	Atenuação Ótica	20
BAPP	Base Avançada de Pesquisa do Pantanal	2
BC	Black Carbon	20
DataONE	Data Observation Network for Earth	11
DER	Entity Relationship Diagram	16
HTML	Hypertext Markup Language	27
OMG	Object Management Group.....	15
PGFA	Programa de Pós Graduação em Física Ambiental	22
PHP	Hypertext Preprocessor	27
RPPN	Reserva Particular do Patrimônio Natural	19
SESC	Serviço Social do Comércio.....	19
SGBD	Sistema Gerenciador de Banco de Dados.....	13
SQL	Structure Query Language	29
UFMT	Universidade Federal de Mato Grosso.....	26
UML	Unified Modeling Language.....	15
URL	Uniform Resource Locator.....	28

RESUMO

KUHNEN, I. A. ESTRATÉGIA COMPUTACIONAL PARA REPRODUTIBILIDADE E ARMAZENAMENTO DE DADOS AMBIENTAIS. Cuiabá, 2020, 58p. Tese (Doutorado em Física Ambiental) - Instituto de Física, Universidade Federal de Mato Grosso.

Experimentos científicos são a base para a validação e a confiabilidade das descobertas científicas e a confiança nos dados fornecidos para estudos é primordial para a garantia da qualidade das informações geradas. Atualmente os processos de coleta, armazenamento e compartilhamento dos dados em grande parte das estações que sustentam os estudos ambientais são realizados de forma manual, comprometendo até mesmo a integridade dos dados gerados. Este trabalho tem por objetivo geral desenvolver uma ferramenta computacional, baseada na gestão, armazenamento e compartilhamento de dados ambientais, para apoiar a reprodutibilidade e a reutilização dos mesmos. A pesquisa foi desenvolvida na Base Avançada de Pesquisas do Pantanal-BAAP, na cidade de Poconé - MT. A metodologia englobou a investigação das características estruturais dos dados gerados pelos equipamentos bem como dos processos que descrevem o ciclo de vida dos dados, de forma a identificar as demandas para o novo sistema e dar subsídios para a elaboração dos diagramas de fluxo de dados, diagramas de entidade e relacionamento e diagramas de atividades, necessários para a descrição do novo sistema proposto. Como resultado, uma ferramenta computacional Web foi desenvolvida, em linguagem PHP, conhecida e de código aberto, em um framework Laravel, para possibilitar a administração dos dados, uma aplicação de um novo processos de ciclo de dados da BAPP, envolvendo as etapas de planejamento, coleta, segurança, descrição, preservação, descoberta, integração, análise e, novamente, o planejamento, possibilitando que o mesmo conjunto de dados gerem diferentes informações e conhecimento. Usado um sistema gerenciador de banco de dados, MYSQL, também de código aberto onde foi parametrizado de forma a garantir elevada robustez e fácil aplicabilidade na replicação dos dados em máquinas distintas.

Palavras-chave: Dados Ambientais, Banco de dados, Reprodutibilidade, Laravel.

ABSTRACT

KUHNEN, I.A.: Computational strategy for reproducibility and storage of environmental data for the environmental physics program. Cuiabá, 2020, 57p. Tese (Doutorado em Física Ambiental) - Instituto de Física, Universidade Federal de Mato Grosso.

Scientific experiments are the basis for the validation and reliability of scientific discoveries and confidence in the data provided for studies is paramount for ensuring the quality of the information generated. Currently, data collection, storage and sharing processes in most stations that support environmental studies are carried out manually, compromising the integrity of the data generated. The general objective of this work is to develop a computational tool, based on the management, storage and sharing of environmental data, to support their reproducibility and reuse. The research was developed at the Advanced Research Base of Pantanal-BAAP, in the city of Poconé - MT. The methodology included the investigation of the structural characteristics of the data generated by the equipment as well as of the processes that describe the data life cycle, in order to identify the demands for the new system and provide subsidies for the elaboration of data flow diagrams, diagrams of entity and relationship and activity diagrams, necessary for the description of the proposed new system. As a result, a web computing tool was developed, in PHP, known and open source, in a Laravel framework, to enable data management, an application of a new BAPP data cycle process, involving the planning steps, collection, security, description, preservation, discovery, integration, analysis and, again, planning, enabling the same set of data to generate different information and knowledge. A database management system, MYSQL, also open source was used where it was parameterized in order to guarantee high robustness and easy applicability in the replication of data on different machines.

Keywords: Ambiental data, Database, Reproducibility, Laravel.

Capítulo 1

INTRODUÇÃO

1.1 PROBLEMÁTICA

Experimentos científicos são a base para a validação e a confiabilidade das descobertas científicas, e a confiança nos dados fornecidos para estudos é primordial para a garantia da qualidade das informações geradas. É essencial que o reuso de dados publicados para a geração de novas análises respeite alguns critérios pois os dados são a base da informação e do conhecimento, mas dados incorretos podem levar a mal-entendidos, que apoiarão decisões erradas.

O gerenciamento de dados ambientais está em constante evolução juntamente com os desafios de lidar com esses dados, as iniciativas para a gestão também foram intensificadas devido ao aumento da complexidade dos processos e análise dos dados. Estes tipos de dados demandam técnicas específicas de análise, pois devem seguir a ordem temporal das observações, pois dificultam a remoção de tendências as variações sazonal ou cíclicas, apresenta complicações na seleção de modelos e na interpretação, além de apresentar dificuldades em lidar com observações perdidas e dados discrepantes.

Destaca-se que a disponibilidade dos dados após a publicação dos mesmos diminui 17% ao ano e que a probabilidade de encontrar um endereço de e-mail de trabalho dos autores de um estudo caem 7% ao ano. Portanto, quanto mais antigo o artigo publicado, menos provável é a descoberta e o reuso dos seus dados.

Dados armazenados em diferentes regiões geográficas geram ainda maior complexidade para seu reuso, uma vez que os pesquisadores precisam confiar em tabelas sem conhecer a origem dessas informações, trabalhando com dados já processados, que muitas vezes podem vir acompanhados de erros de interpretação pelo pesquisador que as processou.

Pode-se destacar as pesquisas desenvolvidas pela Universidade Federal de

Mato Grosso - (UFMT) na Base Avançada de Pesquisa do Pantanal (BAPP), localizada na Baía das Pedras, em Mato Grosso (40 km de Poconé), relacionadas ao Bioma Pantanal, altamente explorado por pesquisadores das áreas de física, química, biologia, entre outras, em todo o mundo.

Atualmente os processos de coleta, armazenamento e compartilhamento dos dados das estações presentes na BAPP, assim como em grande parte das estações que sustentam os estudos ambientais, são realizados de forma manual, comprometendo a integridade dos dados gerados.

1.2 JUSTIFICATIVA

A comunidade acadêmica intensificou esforços para aprimorar o gerenciamento dos dados ambientais no intuito de conseguir conservar os dados ao longo do tempo, para aumentar sua visibilidade, economizar tempo e recursos, documentar e apoiar o acesso livre e aberto.

Para garantir que a reprodutibilidade dos dados oriundos das estações de monitoramento ambiental, como as presentes na BAPP, seja realizada de maneira na qual as decisões na geração do conhecimento sejam confiáveis, é necessário que seja realizado o compartilhamento dos dados brutos, sem nenhum processamento.

O crescente avanço na tecnologia e na ciência tem enfrentado desafios em lidar com dados, principalmente nos processos de coleta e processamento de dados científicos. Tal fenômeno fez desenvolver no ramo da ciência a chamada *e-science*, pesquisas que têm necessidade de lidar com grandes volumes de dados, utilizando-se de métodos computacionais sofisticados e computação de alto desempenho. Tais pesquisas envolvem múltiplas fontes de dados e diversas atividades realizados por pesquisadores em vários lugares do mundo e utilizam-se de sistemas capazes de disponibilizar estes dados e publicações para difundir a discussão da ciência.

1.3 OBJETIVO GERAL

Este trabalho tem por objetivo geral desenvolver uma ferramenta computacional, baseada na gestão, armazenamento e compartilhamento de dados ambientais, para apoiar a reprodutibilidade e a reutilização dos mesmos.

1.4 OBJETIVOS ESPECÍFICOS

Para alcançar tal objetivo foi necessária a realização dos seguintes objetivos específicos:

- Analisar os processos e formatação dos arquivos oriundos das estações de monitoramento ambiental sobre o Pantanal Mato-grossense;
- Desenvolver diagramas para a compreensão e desenvolvimento do sistema;
- Desenvolver uma ferramenta computacional web para armazenamento e gestão de dados;
- Preparar uma ferramenta para replicabilidade dos dados em banco de dados distintos;

Capítulo 2

FUNDAMENTAÇÃO TEÓRICA

2.1 DADOS AMBIENTAIS

Segundo Maddala e Lahiri (1992), em estudos ambientais deve ser considerada a interdependência entre observações sequenciais no tempo, sendo este comportamento atribuído a séries temporais. O estudo deste conjunto de dados tem o propósito de compreender e descrever a dinâmica e a estrutura temporal dos mesmos, podendo compreender uma sequência única de dados (univariada), ou de várias coleções de dados (multivariadas) (GUJARATI D. N.; PORTER, 2011).

Pode-se destacar que estes tipos de dados demandam técnicas específicas de análise, devem seguir a ordem temporal das observações, dificultam remoção e estimativa de tendências e variações sazonal ou cíclicas, apresenta complicações na seleção de modelos e na interpretação das ferramentas, além de apresentar dificuldades em lidar com observações perdidas e dados discrepantes (EHLERS, 2007).

A análise de uma série temporal, segundo Cardoso (2001), deve iniciar-se com a modelagem do fenômeno estudado, seguido do comportamento da série, na qual é possível fazer-se estimativas, concluindo com a avaliação dos fatores que direcionaram o comportamento da série, estabelecendo relação de causa e efeito entre duas ou mais séries.

Kuhnen (2016), Oliveira (2015) destacam que, para a garantia da qualidade das análises de séries temporais, como em conjunto de dados ambientais, é necessário que o processamento dos dados seja realizada por um especialista para guiar essas operações. Moresi et al. (2010) ilustram o ciclo desses dados no contexto do gerenciamento ambiental, vide a figura 1.

A primeira etapa engloba o planejamento das atividades a serem realiza-

das, como a definição dos dados a serem coletados e o período de coleta, sendo determinante para que os questionamentos feitos nas pesquisas sejam devidamente respondidos. A segunda etapa envolve o armazenamento e processamento dos dados que possibilitam a interpretação dos mesmos na etapa seguinte. Nesta etapa deve-se ter o cuidado na garantia dos dados armazenados, para que representem corretamente os dados coletados. Na sequência os dados são organizados para que sejam extraídas informações que orientem as tomadas de decisões. A etapa seguinte envolve a análise das informações extraídas, é neste momento que as hipóteses são testadas e validadas. Por último têm-se a etapa de disseminação na qual ocorre a aplicação dos modelos já validados, vide figura 1.



Figura 1: Fluxo de atividades no monitoramento ambiental.
 FONTE: Adaptado de Moresi et al. (2010).

Este fluxo de atividades pode ser aplicado a demais áreas e padrões de dados dentro da abordagem específica da gestão de dados. A maior dificuldade no monitoramento de informações ambientais está na recepção, armazenamento e processamento dessas informações. Os dados das estações ambientais são, na maioria das vezes, coletados manualmente na estação, ou enviados para um servidor. No primeiro caso um usuário é responsável pelas operações de armazenamento, tratamento, organização e disponibilização dos dados, tornando-os dependentes da capacidade daquele, prejudicando a utilização dos dados por pesquisadores de outras áreas (COUTINHO et al., 2015).

2.2 GESTÃO DOS DADOS

A gestão de dados científicos acompanha todo o ciclo de vida dos dados, iniciando com a coleta, passando por processamentos de limpeza, curadoria, anotação, indexação e transformação. Segundo Medeiros (2018), grande parte da pesquisa científica de hoje exige análise de processamento de dados, e com isto, o planejamento da gestão dos dados passou a fazer parte integral da metodologia científica.

É importante definir dados, como informação e conhecimento, que são relacionados, mas que possuem diferentes definições. São representações de fatos na forma de textos, números, gráficos e imagens (BRACKETT; EARLEY, 2009).

A informação são os dados em um contexto. Sem um contexto os dados não possuem sentido. A partir da interpretação do contexto relacionado aos dados cria-se uma informação. Segundo Brackett e Earley (2009), este contexto deve incluir o significado dos elementos e dos termos, o formato e o período representado pelos dados, os objetivos que definiram a geração dos dados e a relevância dos mesmos para um determinado uso.

O conhecimento é a informação em uma perspectiva, ligado a um ponto de vista, tendo como base a interpretação de padrões, como tendências, oriundas de outras informações e experiências. O conhecimento é obtido quando o significado das informações é compreendido (BRACKETT; EARLEY, 2009). Vide a figura 2.

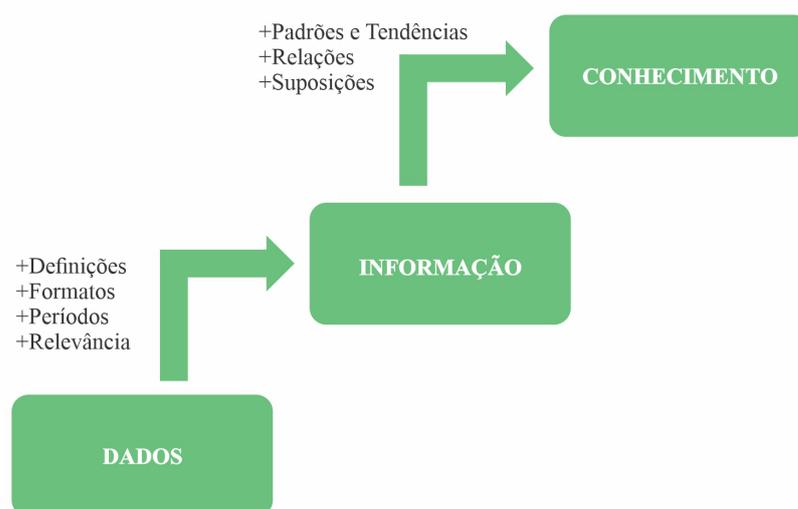


Figura 2: Dados, informação e conhecimento
FONTE: Adaptado de Brackett e Earley (2009).

O gerenciamento de dados é responsável por definir, planejar, implementar e executar as estratégias, procedimentos e práticas necessárias para gerenciar efetivamente os recursos de dados e informações da organização, incluindo sua definição, padronização, organização, planos de proteção e uso (BRACKETT; EARLEY, 2009). As atividades de gerenciamento de dados foram consideradas em diferentes áreas da organização, estendendo-se a parceiros, fornecedores e consumidores.

No nível institucional, envolve gerentes que usam dados para tomar decisões e profissionais de nível operacional responsáveis por coletar, gerar e analisar dados. Entre os principais objetivos da gestão de dados destacam-se a garantia da qualidade, correta utilização e reutilização, decisões ágeis, confiabilidade, segurança e gestão de riscos e por último a manutenção a longo prazo (BRACKETT; EARLEY, 2009) e (STRASSER et al., 2012).

2.2.1 Gestão de Dados Científicos

Gerenciamento de dados é um conjunto de várias práticas aplicadas ao processo de pesquisa que ajudam a descoberta, compreensão, reutilização e menor probabilidade de perda ou esquecimento dos dados (BRINEY, 2015).

Segundo Borgman (2012), ao tornar os dados visíveis para outros pesquisadores, por meio de um repositório reconhecido, pode-se provar a importância da pesquisa e, ao mesmo tempo, garantir que as informações possam ser usadas para outros fins. Economizar tempo a médio e longo prazo através do planejamento da gestão dos dados.

A disponibilização dos dados, permite que outros pesquisadores encontrem, entendam e reutilizem os dados, sem necessidade de intermédio do pesquisador que as gerou. Ao armazenar os dados em um repositório, a segurança aumenta, assim o investimento de tempo e recurso são preservados.

Borgman (2012) cita ainda que novas pesquisas podem ser realizadas com base nos dados existentes, sem necessidades de coletar dados adicionais para processar os dados. Possibilitam que os outros pesquisadores possam avançar em pesquisas inacabadas ou realizar descobertas inesperadas, compartilhar o conhecimento da comunidade científica junto aos pesquisadores para expandir suas pesquisas.

No entanto, desafios atuais ainda dificultam a aplicação das boas práticas de gestão de dados, como o armazenamento de dados em depósitos isolados, criados por organizações, seus setores e até pelos próprios pesquisadores e colaboradores. Embora os maiores problemas sejam causados por silos de dados

isolados, que não permitem acesso a dados, outro desafio é gerado pelas próprias iniciativas das comunidades científicas, que geram grandes bancos de dados de tipos de dados específicos. Esses bancos de dados têm pouca ou nenhuma integração com bancos de dados direcionados a outros tipos de dados (PETERSON et al., 2015). Com isso, o pesquisador ainda tem um grande esforço no processo de obtenção dos dados necessários para sua pesquisa.

Outro detalhe são dados órfãos, onde as informações são perdidas em uma instituição em situações como a aposentadoria de um pesquisador, o roubo ou a destruição de um notebook que não teve seus dados copiados para um servidor seguro ou o simples fato de que informações importantes foram esquecidas em um meio de armazenamento inacessível, como computador, mídia digital ou *pen drive*.

O conhecimento sobre informações é gradualmente perdido após a conclusão de seu objetivo principal, como a publicação dos resultados de uma pesquisa (MICHENER et al., 1997). Assim, sem o uso de boas práticas para gerenciar os dados produzidos durante uma pesquisa, esse conhecimento acaba sendo esquecido ao longo do tempo. A figura 3 ilustra uma representação da entropia dos dados científicos.

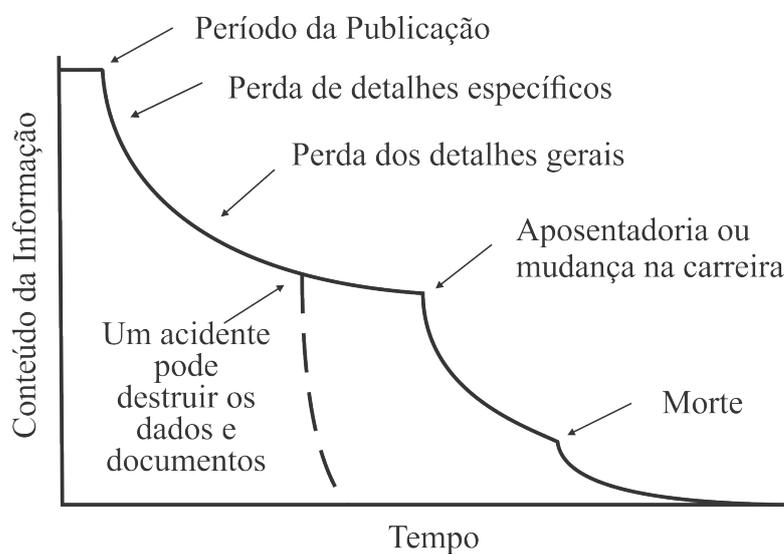


Figura 3: Representação de entropia dos dados científicos.
 FONTE: Adaptado de Michener et al. (1997).

Na forma de análise de cauda longa dos dados, esta em contraste com o volume de dados gerados por grandes projetos científicos, uma quantidade menor de dados, mas variada e distribuída; em paralelo, é gerada por pesquisadores individuais e pequenos grupos de pesquisa (MICHENER et al., 1997). Ao contrário de grandes projetos, que possuem várias equipes de pesquisadores, ferramentas

e procedimentos computacionais modernos para gerenciamento de dados, esses pequenos projetos têm recursos, ferramentas e pesquisadores limitados. Como consequência, os dados gerados nesses projetos geralmente não estão disponíveis na Web e raramente são reutilizados e preservados a longo prazo. Esses pequenos projetos formam a ciência (HEIDORN, 2008; HOWE et al., 2011; WALLIS et al., 2013).

Conforme apresentado em Ferguson et al. (2014), os estudos mostram uma distribuição distorcida entre o tamanho dos conjuntos de dados em relação à diversidade dos conjuntos de dados. Grandes projetos científicos, com seus dados homogêneos e organizados, representam apenas uma pequena parte da diversidade de dados. A grande parte dos dados está na cauda longa, produzindo uma vasta variedade de conjuntos de dados. Na extrema direita da cauda longa estão os dados não publicados ou indisponíveis.

Apesar dos dados gerados pelos projetos que compõem o cauda longa representarem apenas uma pequena parte do volume dos dados científicos totais, estes dados formam a base de diversas teorias científicas (HEIDORN, 2008). Por isso, tornar os dados do cauda longa acessíveis e reutilizáveis é um dos grandes desafios da atualidade (WALLIS et al., 2013). A figura 4 ilustra uma representação da cauda longa dos dados científicos.

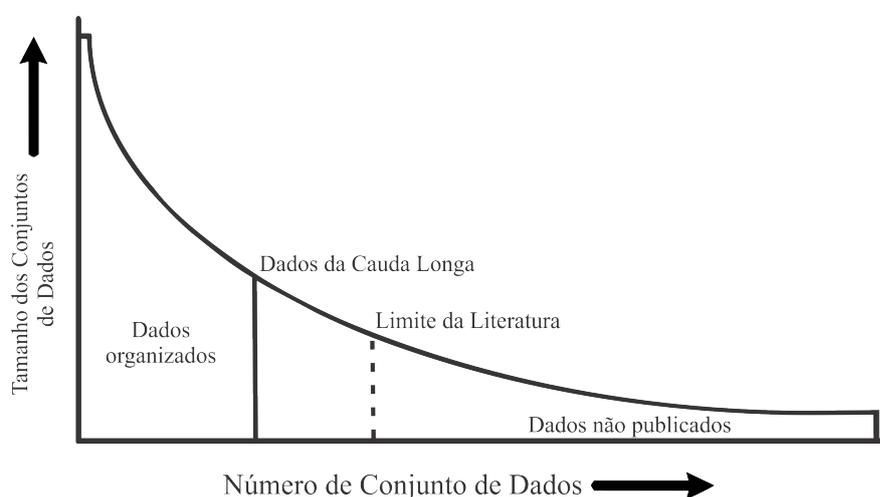


Figura 4: Esquema apresentando a cauda longa dos dados científicos.
 FONTE: Adaptado de Michener et al. (1997).

Sempre há necessidade de análise, em todo ciclo de vida dos dados, desde a recepção, validação, análise, compartilhamento e seu armazenamento. Sem essas melhorias, os dados continuarão sendo perdidos, gerando custos devido a necessidade de repetição de coletas. Em vários casos a perda é permanente (LYNCH, 2008).

2.2.2 Ciclo de Vida dos Dados

Para um gerenciamento de dados eficiente é importante entender os vários estágios pelos quais os dados passam durante o processo científico. Essa sequência de etapas é chamada de ciclo de vida dos dados. O uso desse ciclo é comum no gerenciamento de dados, pois ajuda a identificar o papel dos dados em cada uma das etapas de um processo, além de permitir a definição das melhores estratégias, ferramentas e boas práticas para o gerenciamento de dados ao longo de seu ciclo de vida (BRACKETT; EARLEY, 2009) ; (RÜEGG et al., 2014).

Com o desenvolvimento da ciência da computação, a importância dos dados na pesquisa aumentou e o ciclo de vida tradicional dos dados científicos está sendo substituído por um novo ciclo de vida.

2.2.2.1 O Ciclo de Vida dos Dados Tradicional

Este ciclo de vida existe desde que a publicação de artigos científicos tornou-se um padrão (BRINEY, 2015). O ciclo inicia com o planejamento, aquisição, e termina com a publicação dos resultados da pesquisa. Este ciclo de vida é apresentado na figura 5.



Figura 5: Representação do antigo ciclo de vida dos dados.
FONTE: Adaptado de Briney (2015).

Nesse processo os dados desempenham um papel importante na fase de coleta e análise, que é uma atividade orientada a dados. Em menor grau, no entanto, os dados também são incluídos nas fases de planejamento e publicação do projeto. Esse ciclo de vida mostra claramente que os dados são usados como um meio e não como um fim.

O objetivo do processo é a publicação de um artigo científico. Os dados são considerados subprodutos da pesquisa e geralmente são perdidos ou esquecidos após o término do projeto ou a publicação do artigo científico (BRINEY, 2015; VINES et al., 2014). Um estudo realizado por Vines et al. (2014) identificou que a disponibilidade dos dados após a publicação diminui 17% ao ano e que a

probabilidade de encontrar um endereço de e-mail de trabalho dos autores de um estudo caem 7% ao ano. Portanto, quanto mais antigo o artigo publicado, menos provável é a descoberta e o reuso dos seus dados.

Esse é um dos maiores problemas no atual ciclo de vida dos dados, pois os dados da pesquisa, em geral, têm um valor além da pesquisa original, tanto no laboratório quanto na instituição que conduziu a pesquisa e fora dela. No entanto, devido ao processo de pesquisa tradicional que não valoriza a reutilização de dados, há pouco incentivo para a manutenção e preservação desses dados após sua publicação.

Embora esse processo de trabalho funcione há centenas de anos, os avanços na ciência da computação mostraram a importância dos dados e sua preservação e reutilização. Gerando a necessidade de alterar o ciclo de vida dos dados científicos.

2.2.2.2 O Novo Ciclo de Vida dos Dados

O novo ciclo contém etapas com foco no compartilhamento, na preservação e no reuso dos dados gerados no processo de pesquisa. Diversas instituições criaram seus próprios modelos de ciclo de vida dos dados, mas, em geral, seguem uma estrutura semelhante, voltada ao compartilhamento e reuso. Na figura 6 é apresentado o modelo de ciclo de vida dos dados definido pelo *Data Observation Network for Earth* (DataONE).

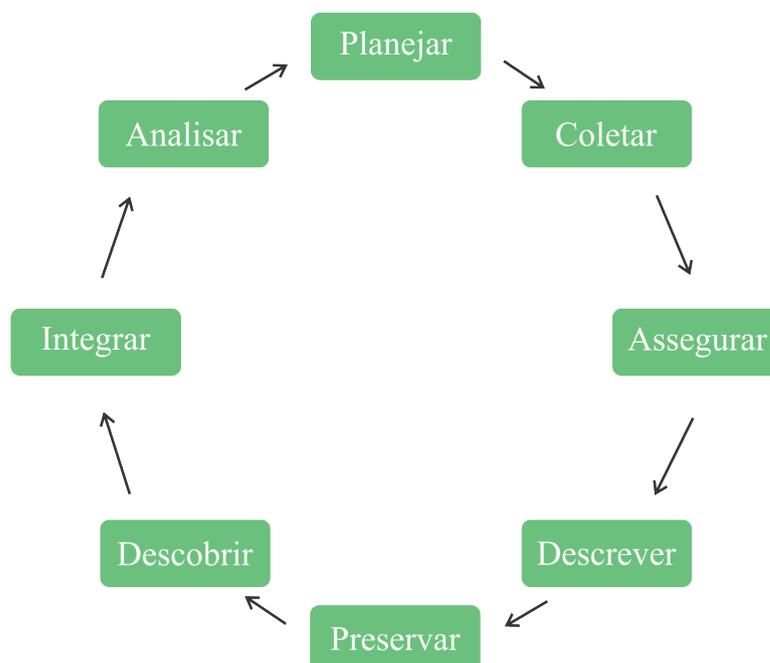


Figura 6: Representação do novo ciclo de vida dos dados.
 FONTE: Adaptado de Strasser et al. (2012).

Outro modelo de ciclo de vida, demonstra de forma explícita o reuso dos dados, foi criado por Michener et al. (2011) e é apresentado na figura 7. O primeiro ciclo inicia com a definição da pergunta de pesquisa, passando pela aquisição dos dados, a aplicação dos processos de qualidade, as análises e, finalmente, a publicação, que também produz novas perguntas de pesquisa, iniciando o ciclo novamente. O segundo ciclo surge a partir dos dados que foram preservados e compartilhados.

Neste ciclo, novas perguntas de pesquisa são respondidas a partir do reuso dos dados existentes, que são integrados e analisados, gerando novas descobertas científicas, que também tem seus dados preservados. Este ciclo, baseado no reuso de dados, é mais ágil e menos custoso, impulsionando as novas descobertas científicas.

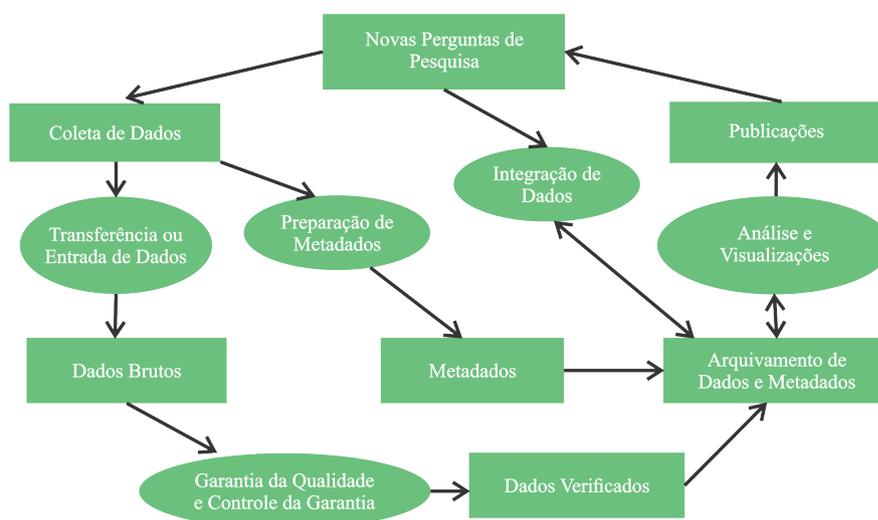


Figura 7: Representação do ciclo de vida dos dados científicos.
 FONTE: Adaptado de Michener et al. (2011).

O gerenciamento de dados científicos requer esforços nas diferentes etapas do ciclo de vida dos dados. Se algumas dessas etapas forem negligenciadas, as etapas posteriores não poderão gerar resultados eficientes. Portanto, para a geração de conhecimento científico reproduzível e reutilizável, os esforços devem começar a partir do planejamento da pesquisa, através das várias etapas intermediárias, até a publicação, preservação e compartilhamento de dados.

2.3 BANCO DE DADOS

Um banco de dados é uma coleção estruturada e persistente de informações sobre algum aspecto do mundo real, organizado e armazenado de uma maneira

que facilita recuperação e modificação eficientes (REZENDE, 2006). A estrutura de um banco de dados é determinada por um resumo do modelo de dados, que possibilita diferenciá-lo de um arquivo de dados. O modelo de dados mais popular é o relacional que representa dados como um conjunto de tabelas. O banco de dados inclui instâncias de entidades e os relacionamentos entre elas. Por exemplo: um banco de dados para uma escola pode conter instâncias de estudantes, professores, cursos, quais alunos fazem quais cursos, quais professores ensinam quais cursos (PERIS, 2018).

Considerando a existência de um programa que gerencia esses dados, obtêm-se um Sistema Gerenciador de Banco de Dados (SGBD), que simplifica a criação e manutenção de um Banco (DATE, 2004). O SGBD garante a segurança dos dados armazenados além de permitir operações como: inserir, alterar, excluir, obter e atualizar dados em um sistema (ELMASRI et al., 2005).

2.3.1 Replicação de Dados

Replicação de dados é o processo para realização de cópias de uma coleção de dados, onde cada item possui várias cópias físicas, cada uma delas localizada em um nó diferente, também conhecido como *site* (normalmente uma máquina física). Dependendo do contexto e do tipo de arquitetura de replicação, o termo réplica pode se referir a um dos fatores físicos copiados de um item de dados específico ou para um site com todas as suas cópias de dados.

A replicação de dados pode servir para propósitos diferentes, em primeiro lugar, pode ser usado para aumentar a disponibilidade e fornecer tolerância a falhas, já que os dados podem, em princípio, ser acessados desde que uma réplica esteja disponível. Em segundo lugar, pode fornecer acelerado tempo de resposta a operação (PERIS, 2018).

A replicação pode ser configurada em um servidor, atuando como principal e um ou mais servidores atuando como secundários. O principal grava em um arquivo de registro as alterações de todos os comandos de atualizações da base de dados. Desta forma, todas as alterações ocorridas no principal são imediatamente replicadas para os outros servidores secundário.

Observa-se uma série de vantagens do uso da replicação, tais como:

- Equilíbrio de carga: a replicação, geralmente não precisa de uma largura de banda muito intensiva, pode iniciar e parar conforme a sua vontade e o servidor secundário pode trabalhar distante do principal.
- Backup: a replicação é uma técnica valiosa para ajudar com *backups*.

- Alta disponibilidade: se um servidor principal falhar têm-se outro servidor para supri-lo.

A figura 8 a seguir mostra com mais detalhes sobre a replicação de dados:

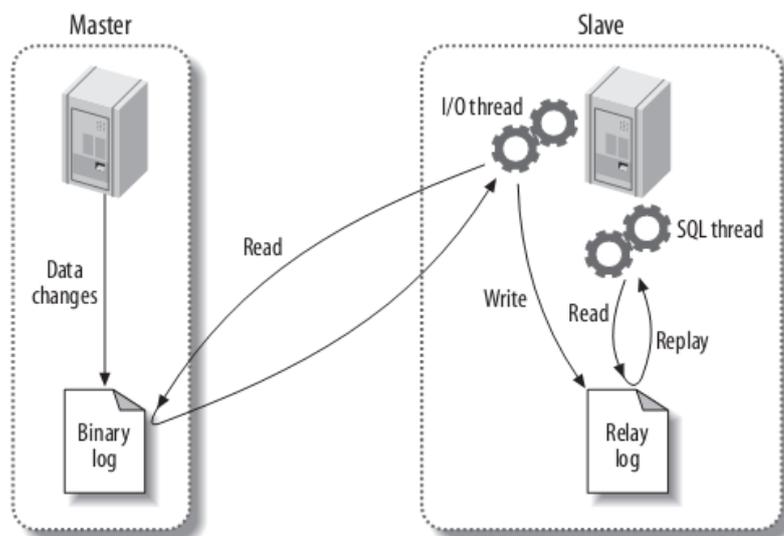


Figura 8: Replicação de dados.
 FONTE: Adaptado de Viana (2011)

2.4 LINGUAGEM DE MODELAGEM PADRONIZADA (*Unified Modeling Language-UML*)

A UML é uma linguagem de uso geral na Engenharia de Software, no domínio de programas orientados a objetos. A padronização é gerenciada, e foi criada, pelo *Object Management Group* (OMG) e foi o primeiro padrão adicionado à lista de tecnologias adotadas do OMG. Desde então tornou-se o padrão da indústria para modelagem de programas (SPECIFICATION, 2007). A UML oferece uma forma padrão para visualizar a arquitetura do sistema, incluindo elementos como: atividades, atores, processos de negócio, esquemas de banco de dados, componentes, comandos de linguagens de programação e componentes de programas reutilizáveis.

É importante deixar bem clara a diferença entre modelo UML e o conjunto de diagramas de um sistema. Um diagrama é uma representação gráfica (parcial) de um modelo do sistema. O modelo contém, também, a documentação que orienta os seus elementos e diagramas (como os Casos de Uso escritos). Os diagramas UML representam duas diferentes visões de um modelo de sistema (HOLT, 2004):

- Estática (ou Estrutural): destaca a estrutura estática do sistema, utilizando objetos, atributos, operações e relacionamentos. Inclui os Diagramas de Classes, Componentes e Pacotes, por exemplo.
- Dinâmica (ou Comportamental): destaca o comportamento dinâmico do sistema, mostrando as colaborações entre objetos e as mudanças internas nos objetos. Inclui os Diagramas de Sequência, Atividades e Máquina de Estado, por exemplo.

2.4.1 Diagrama de Entidade Relacionamento (DER)

Um diagrama entidade relacionamento (DER) é um fluxograma que representa entidades, tais como pessoas, objetos ou conceitos, estão relacionados/interligados em um sistema. Diagramas DER são utilizados para projetar bancos de dados relacionais, sistemas de informações empresariais, educação e pesquisa. Utilizam um conjunto definido de símbolos, como retângulos, diamantes, ovais e linhas de conexão que representam a interconectividade de entidades, relacionamentos e seus atributos (LUCIDCHART, 2020).

As linhas de conexão (cardinalidade) define a relação entre duas entidades ou conjunto de entidades, e podem ser representadas como ilustrada na figura 9.

 Zero ou um	 Vários	 Um
 Um (ou somente um)	 Zero ou Vários	 Um ou Vários

Figura 9: Simbologia das conexões do DER.

2.4.2 Diagrama de Atividades

São modelos para representar tanto a estrutura quanto o comportamento de um sistema e, com base neles, construir e programar o modelo executável, que é o sistema materializado. A estrutura representa os elementos estáveis que não se alteram na utilização do sistema, enquanto o comportamento representa os elementos dinâmicos, modificados pelas ações do sistema ou do usuário (VENTURA, 2020).

Diagramas de atividade possibilitam a demonstração da lógica de um algoritmo, descrevem as etapas realizadas em um caso de uso UML, ilustram um processo de negócio ou fluxo de trabalho entre usuários e o sistema, simplificam e melhoram qualquer processo ao esclarecer casos de uso complicados e modelam elementos de arquitetura de programa, como método, função e operação.

Alguns dos componentes mais comuns de um diagrama de atividade estão ilustrados na figura 10.

 <p>Representa o começo de um processo</p>	 <p>Indica as atividades que compõem um processo modelado</p>	 <p>Mostra a direção do fluxo</p>
 <p>Símbolo da decisão (com dois caminhos ramificados e com texto de condição)</p>	<p>[Condition]</p> <p>Símbolo de condição (é colocado ao lado do marcador para avisar qual condição uma atividade deve separar)</p>	 <p>Representa o estado final de uma atividade</p>

Figura 10: Simbologia das conexões do Diagrama de Atividades.

2.4.3 Diagrama de Fluxo de Dados (DFD)

São modelos que ilustram o fluxo e a transformação de dados para um determinado processamento. É uma representação visual de como os dados fluem através de um sistema, possibilitando a visão clara de onde os dados vem, para onde vão, e como serão armazenados (LUCIDCHART, 2020).

Alguns dos componentes mais comuns de um diagrama de fluxo de dados estão ilustrados na figura 11.

 <p>Uma atividade onde a manipulação e a transformação de dados ocorrem</p>	 <p>Como os dados fluem</p>
 <p>Pode ser uma pessoa, sistema ou aplicação, onde começam ou terminam</p>	 <p>Os dados necessários ou produzidos nos processos são armazenados</p>

Figura 11: Simbologia das conexões do DFD.

Capítulo 3

MATERIAL E MÉTODOS

Neste capítulo são apresentados os locais de origem dos dados de estudo, bem como as características dos mesmos e os equipamentos utilizados em cada um deles. É descrito também o processo de desenvolvimento do ambiente proposto, as linguagens utilizadas e a definição da arquitetura.

3.1 LEVANTAMENTO AMBIENTAL SOBRE O PANTANAL MATO-GROSSENSE

3.1.1 Local de Origem dos Estudos

Os dados utilizados neste trabalho foram retirados da BAPP, unidade que pertence à Reserva Particular do Patrimônio Natural (RPPN) do Serviço Social do Comércio (SESC) - Pantanal, localizada na cidade de Poconé - MT, na figura 12.

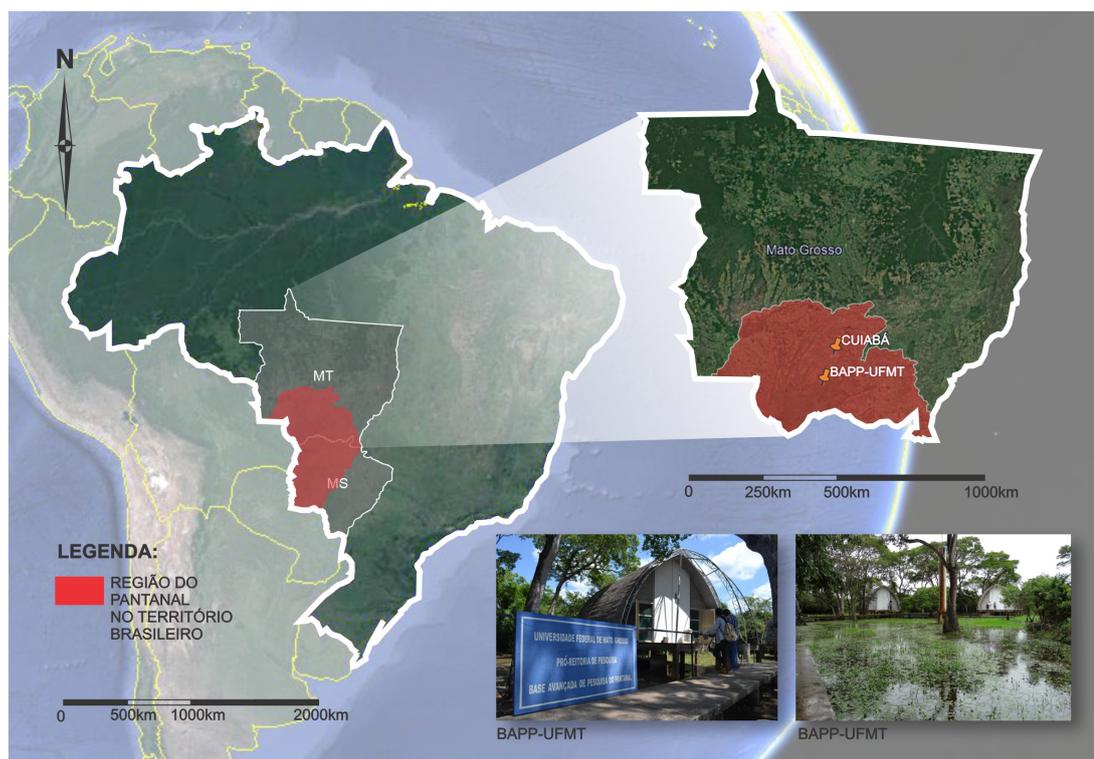


Figura 12: Região do pantanal matogrossense onde encontra-se a BAPP-UFMT, local onde encontra-se os equipamentos da retirada dos dados para desenvolvimento do sistema.

Na BAAP são realizados monitoramentos de dados ambientais envolvendo medidas de propriedades óticas de aerossóis, micrometeorologia e fluxos de energia e matéria. Diante da diversidade de dados a seleção de uma amostra que possibilite a definição de características que tornassem o sistema proposto apto a gerir todos os demais dados. A amostra foi concebida a partir da necessidade de uma definição da proporcionalidade de colunas e linhas na apresentação dos dados dentro da plataforma computacional, para fins de elaboração do código fonte, foram selecionados os equipamentos que apresentaram maior volume de dados gerados, sendo eles o aetalômetro e o nefelômetro. Finalizado as etapas de teste, os demais equipamentos poderão ser adicionados conforme demanda do PGFA.

3.1.2 Aetalômetro

O aetalômetro, vide figura 13, é um instrumento que fornece uma medição óptica em tempo real de aerossóis de *black carbon* (BC), medida em 7 comprimentos de onda diferentes, entre 370nm a 950nm. O princípio do aetalômetro é medir a atenuação (ATN) de um feixe de luz transmitido através de um fil-

tro, enquanto o filtro coleta continuamente uma amostra de aerossol. A ATN é medido através de um ponto "sensor" do filtro, no qual o aerossol é coletado, e um ponto "de referência" do filtro, como uma verificação da estabilidade da fonte óptica (HOLANDA, 2015; PALACIOS, 2017).



Figura 13: Aetalômetro utilizado na BAPP.

A atenuação ótica é definida como:

$$ATN = 100 \ln\left(\frac{I_0}{I}\right) \quad (1)$$

Onde I_0 representa a intensidade da radiação transmitida pelo filtro limpo, e I representa a intensidade de radiação transmitida após a exposição do filtro.

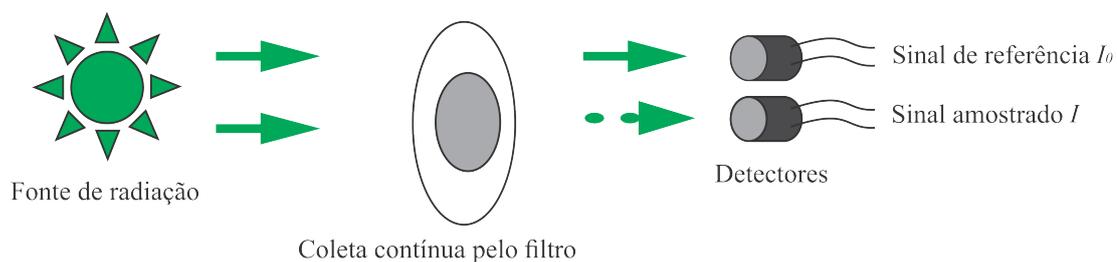


Figura 14: Princípio de funcionamento do aetalômetro.
Fonte: Palacios (2017).

3.1.3 Nefelômetro

O nefelômetro, vide figura 15, é um instrumento que mede o coeficiente de espalhamento da luz do aerossol. Ele detecta propriedades de dispersão medindo a luz dispersa pelo aerossol e subtraindo a luz dispersa pelo gás, por meio de

uma câmera volumétrica de análise. O nefelômetro *Radiance Research* mede um comprimento de onda, a 525 nm. Uma válvula é acionada para escoar toda a amostra de aerossóis do interior do instrumento para que se possa medir o sinal de ar limpo do ambiente. Este sinal é subtraído do original para que o espalhamento medido seja exclusivamente devido à presença de partículas de aerossol (HOLANDA, 2015; PALACIOS, 2017).



Figura 15: Nefelômetro utilizado na BAPP.

3.1.4 Descrição do Processo de Coleta

Por meio de visitas *in loco* e acompanhamento dos processos de coleta realizados pela equipe de instrumentação do PGFA na BAPP, foi construído um diagrama de atividade de forma a possibilitar um diagnóstico dos processos atuais para orientar a construção do sistema, conforme mostrado na figura 16.

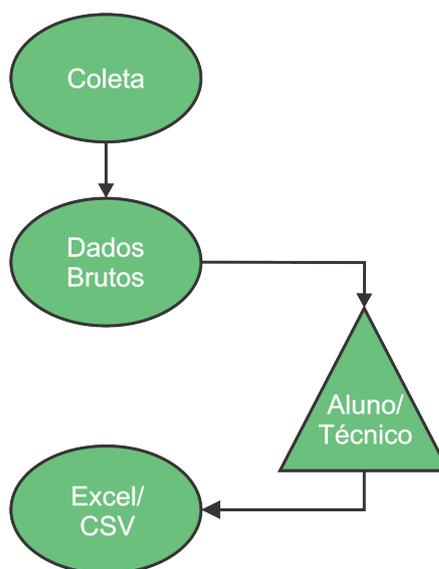


Figura 16: Representação do Processo Antigo do processamento dos dados.

Como pode ser visualizado na figura 16 a coleta é realizada manualmente por um aluno ou técnico que se desloca até a estação e realiza a gravação dos

dados em arquivos de um equipamento físico no formato de CSV/Excel ¹, dados esses que não apresentam nenhum tipo de alteração ou tratamento prévio. Vale destacar que o deslocamento da UFMT, campus Cuiabá, até a BAPP é de aproximadamente 160 km.

3.2 DIAGRAMAS DO SISTEMA

Tendo como base os processos atuais de gestão dos dados na BAPP-UFMT, foram gerados os diagramas de fluxo de dados, entidade relacionamento e de atividade para descrição dos processos e fluxos que os dados precisam percorrer no sistema proposto, de forma a viabilizar o desenvolvimento do mesmo na plataforma WEB de forma automatizada.

Com o sistema proposto, o usuário passará a armazenar os dados em um banco de dados, facilitando a inserção, manutenção e replicação desses dados, e apresentação de relatório e gráficos para análise dos mesmos, como representado no diagrama de atividade ilustrado na figura 17.

¹CSV é um formato de arquivo que significa *comma separated values* (valores separados por vírgulas).

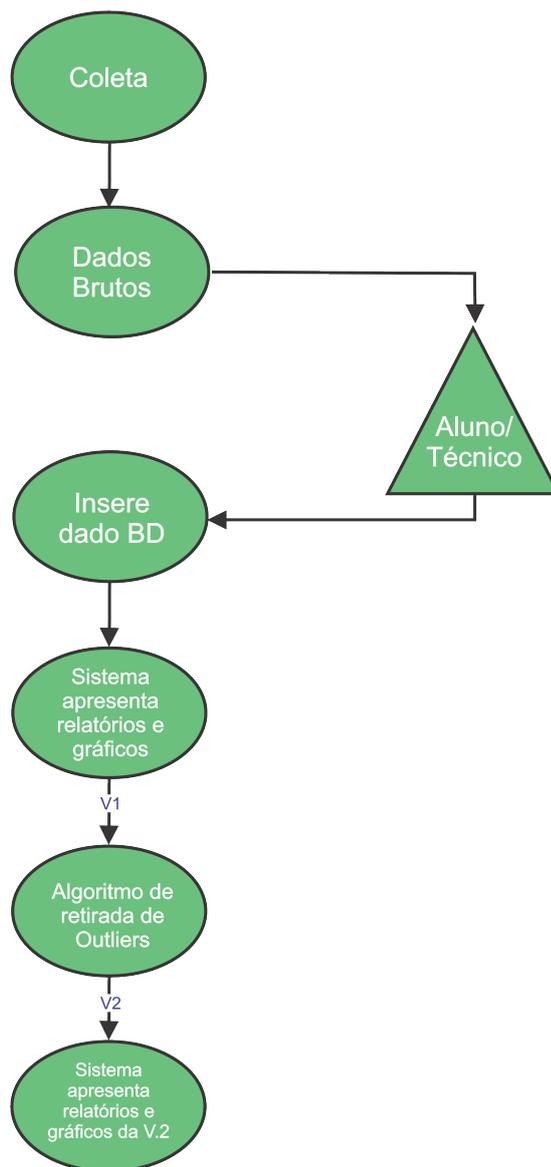


Figura 17: Representação do Processo Automatizado do processamento dos dados.

O banco de dados foi modelado de acordo com as demandas dos sensores selecionados como representativos de todos os demais encontrados na BAPP. O diagrama de entidade relacionamento (figura 18) apresenta a estrutura de tabelas e campos necessária para o desenvolvimento da aplicação.

USERS	
ID	
NAME	
E-MAIL	
E-MAIL-VERIFIED_AT	
PASSWORD	
REMEMBERTOKEN	

NEPHELOMETRO	
ID_NEPHELOMETRO	
DATETIME	
TYPE	
SCAT_COEF	
AIR_TEMP	
CELL_TEMP	
RH	
PRESSURE	

AETALOMETRO	
ID_AETALOMETRO	
DATE	FRACTION_3
TIME	ATENUATION_3
CONC_1	SZ_4
CONC_2	SB_4
CONC_3	RZ_4
CONC_4	RB_4
CONC_5	FRACTION_4
CONC_6	ATENUATION_4
CONC_7	SZ_5
VFLOW	SB_5
SZ_1	RZ_5
SB_1	RB_5
RZ_1	FRACTION_5
RB_1	ATENUATION_5
FRACTION_1	SZ_6
ATENUATION_1	SB_6
SZ_2	RZ_6
SB_2	RB_6
RZ_2	FRACTION_6
RB_2	ATENUATION_6
FRACTION_2	SZ_7
ATENUATION_2	SB_7
SZ_3	RZ_7
SB_3	RB_7
RZ_3	FRACTION_7
RB_3	ATENUATION_7

Figura 18: Diagrama Entidade Relacionamento onde apresenta a estrutura das tabelas criadas no banco de dados.

O diagrama apresenta cada tabela criada no banco de dados, contendo no topo o nome da tabela, em seguida, sua chave primária, e nas linhas seguintes os nomes dos campos. Cada campo é criado pela framework, onde o tipo e tamanho de cada campo é definido na atribuição da *migration* vide a figura 22 onde apresenta a parte de código fonte criada. As cardinalidades apresentam a permissão de acesso dos usuários para cada tabela, podendo ser de Um para Um, Zero para Um, Um para Vários, etc.

3.3 DESENVOLVIMENTO DA FERRAMENTA WEB

3.3.1 Requisitos

Para a construção da arquitetura computacional da ferramenta proposta foram definidos requisitos computacionais, de conteúdo, de gerenciamento e de uso que serão descritos a seguir.

3.3.1.1 Requisitos Computacionais

De forma a possibilitar a coleta automática dos dados, a estrutura computacional deve apresentar acesso a uma rede 3G/4G e um servidor web para

hospedagem do site. Destaca-se que já foi solicitada, para fins deste trabalho, uma hospedagem no site da UFMT, com ajuda da equipe técnica do setor responsável. Quanto ao desenvolvimento da aplicação, a estrutura computacional deve apresentar um *framework* para desenvolvimento, uma linguagem de programação web e um banco de dados relacional.

3.3.1.2 Requisitos de Conteúdo

De forma a possibilitar o entendimento e a consulta dos dados, além da credibilidade e segurança dos mesmos, o sistema deve apresentar os seguintes requisitos de conteúdos:

- Registro dos dados por meio de um aluno/técnico responsável pela qualidade;
- Registro de um arquivo de log contendo de todas as operações realizadas pelos usuários;
- Registro de usuários para consulta e análise dos dados apresentados através de relatórios e gráficos;
- Registro de autoria e licença de uso dos recursos, juntamente com as regras para reutilização por terceiros;
- Impossibilidade de alteração dos dados uma vez que inseridos no sistema.

3.3.1.3 Requisitos de Gerenciamento

De forma a possibilitar a reprodutibilidade de um experimento, a recuperação e utilização na importação de outros sistemas, além do registro de usuários e suas ações o sistema deve apresentar os seguintes requisitos de gerenciamento:

- Manutenção dos dados e registro dos mesmos durante a execução deste processo;
- Gestão e compartilhamento dos dados de forma aberta;
- Autenticação dos usuários responsáveis pelos registros de dados;
- Exportação dos dados.

3.3.1.4 Requisitos de Uso

De forma a possibilitar a recuperação e análise dos dados, o sistema deve apresentar os seguintes requisitos de uso:

- Possibilitar que o usuário realize consultas;
- Possibilitar consulta e visualização de todos os dados relacionados a um sensor;
- Possibilitar a consulta e recuperação dos dados armazenados no repositório.

3.3.2 Ferramentas de Programação

De forma a atender os requisitos para desenvolvimento do sistema, este tópico apresenta as ferramentas selecionadas devidamente justificadas. É importante destacar que, caso a metodologia seja reproduzida por outros pesquisadores, tais ferramentas podem ser alteradas sem prejuízo da qualidade final.

3.3.2.1 Editor de Código Fonte

Para edição do código fonte foi utilizado o *Visual Studio Code*, na versão 1.41.1, desenvolvida pela *Microsoft* para vários ambientes operacionais e de acesso livre. A escolha se deve ao fato de o mesmo suportar variabilidade de linguagens de programação e um conjunto de recursos que podem estar disponíveis para agilizar seu uso. Além disso, uma documentação completa está disponível para auxiliar no momento da programação.

3.3.2.2 Linguagem de Programação

Para realização da programação foi utilizada a linguagem *Hypertext Pre-processor* (PHP), na versão 7.3, que se trata de uma linguagem de *script open source* de uso geral, muito utilizada, e especialmente adequada para o desenvolvimento web, que pode ser embutida dentro do *Hypertext Markup Language* (HTML). Destaca-se ainda que a mesma se trata de linguagem simples que facilita a manutenção e ampliação do sistema. Outro fator importante é a pluralidade de sistemas operacionais e navegadores de que nada impede o uso do sistema nessa linguagem.

3.3.2.3 Framework

De forma a possibilitar um desenvolvimento rápido em linguagem PHP foi selecionado o *framework* Laravel, na versão 7.2, que possui programa livre e de código aberto. Alguns fatores foram determinantes na escolha, sendo eles:

- Possibilita a utilização de um gerenciador de pacotes ao nível de aplicativo (*Composer*) que facilita a instalação de dependências;
- Apresenta documentação extensa e intuitiva;
- Mapeamento da (URL) digitada no navegador;
- Possui sistema de modelos Blade², que melhora a flexibilidade, não restringe o uso do PHP e reduz o uso do código. Usa a herança e as seções;
- Representação de tabelas no código fonte através de uma classe *Model* que é usada para interagir com essa tabela (*Eloquent e Query Builder*). Os *Models* permitem a consulta de dados em suas tabelas, bem como, operações de *insert, update e delete*;
- Fornecimento de comandos (*Artisan Console*) para facilitar o desenvolvimento da aplicação.

A tabela 1 apresenta os itens mais utilizados no console do *artisan*.

Tabela 1: Comandos utilizados no desenvolvimento.

Comando	Objetivo
make:auth	Exibições básicas de login e registro de telas e rotas
make:controller	Criação da classe de controle
make:mail	Criação da classe de e-mail
make:migration	Criação de um arquivo de migração
make:model	Criação de uma classe de <i>Model</i>
migrate:fresh	Elimina todas as tabelas e execute novamente todas as migrações
migrate:install	Cria o repositório de migração
migrate:refresh	Redefini e executa novamente todas as migrações
migrate:reset	Retrocede todas as migrações de banco de dados
migrate:rollback	Retrocede a última migração do banco de dados
migrate:status	Mostra o status de cada migração

²<<https://laravel.com/docs/7.x/blade>>

3.3.2.4 Sistema Gerenciador de Banco de Dados

O Sistema escolhido foi o *MySQL*, na versão 8.0, por se tratar de um sistema gerenciador de banco de dados relacional de código aberto usado na maioria das aplicações gratuitas, para gerir suas bases de dados. O sistema utiliza a linguagem *Structure Query Language* (SQL), que é a linguagem mais popular para inserir, acessar e gerenciar o conteúdo armazenado em um banco de dados.

3.4 REPLICABILIDADE DOS DADOS

A replicação dos dados foi configurada com o SGBD do Mysql, pois permite configurar um ou mais servidores como réplicas de outro servidor, mantendo seus dados sincronizados com a cópia principal. Em um nível alto, a replicação é um processo simples de três partes:

- O mestre registra as alterações nos dados no registro binário;
- A réplica copia os eventos de registro binário do mestre para o registro de retransmissão;
- A réplica replica os eventos no registro de retransmissão, aplicando as alterações em seus próprios dados;

Capítulo 4

RESULTADOS

Neste capítulo são apresentados as definições do desenvolvimentos da aplicação, e todos os componentes necessários para exibir os dados gerados pelos sensores. A estrutura do capítulo está estruturada de forma a ilustrar como foi preparado o ambiente computacional com toda sua estrutura de pastas até as telas com suas funções.

4.1 DESENVOLVIMENTO DO SISTEMA

O sistema de rotas do Laravel permite registrar método para todos os caminhos http disponíveis, sendo 2 tipos utilizados:

- GET - Envia dados ao servidor, o navegador pega as informações do formulário e coloca junto a URL e envia, separando o endereço da URL.
- POST - Só pode ser gerado através de um formulário *web*, normalmente são utilizados para cadastrar algo no sistema.

O encaminhamento para cada página cria no sistema uma rota, para acesso ao mesmo. Em alguns casos são utilizados formas diretas de acesso, como apresenta a linha 6 a 8 da figura 19. Já em outros casos passa-se o caminho do acesso ao sistema por parâmetro, como apresentado na mesma figura, nas linhas 12 a 17.

```

1  <?php
2
3  use Illuminate\Support\Facades\DB;
4  use Illuminate\Support\Facades\Auth;
5
6  Route::get('/', function () {
7      return view('welcome');
8  });
9
10 Auth::routes();
11
12 Route::get('/nephelometro', 'ControladorNephelometro@index');
13 Route::get('/nephelometro/novo', 'ControladorNephelometro@create');
14 Route::post('/nephelometro', 'ControladorNephelometro@store');
15 Route::get('/nephelometro/apagar/{id}', 'ControladorNephelometro@destroy');
16 Route::get('/nephelometro/editar/{id}', 'ControladorNephelometro@edit');
17 Route::post('/nephelometro/{id}', 'ControladorNephelometro@update');
18
19
20 Route::get('/maap', 'ControladorMaap@index');
21 Route::get('/maap/novo', 'ControladorMaap@create');
22 Route::post('/maap', 'ControladorMaap@store');
23 Route::get('/maap/apagar/{id}', 'ControladorMaap@destroy');
24 Route::get('/maap/editar/{id}', 'ControladorMaap@edit');
25 Route::post('/maap/{id}', 'ControladorMaap@update');

```

Figura 19: Linhas de códigos das rotas do sistema.

Ao realizar a criação das *migrations*, que são responsáveis por um controle de versão do seu banco de dados, permitirão que se modifique e compartilhe facilmente o esquema do banco de dados do sistema. Com isso, independente de qual banco de dados se utiliza no sistema a responsabilidade de se comunicar é do Laravel. Segue abaixo a linha de comando utilizada para criação de uma *migrations*, vide a figura 20.

```
$ php artisan make:migration nephelometro
```

Figura 20: Comando para criação da Migration.
 FONTE: <https://laravel.com/docs/6.x>

Pode-se observar na figura 21 a seguir, algumas *migrations* criadas para o desenvolvimento do sistema, tendo em vista que cada uma dela corresponde a uma tabela no banco de dados.

```

1 2019_04_12_031216_criar_tabela_nephelometro.php
2 2019_04_16_203950_criar_tabela_maap.php
3 2019_04_16_204501_criar_tabela_aethalometro.php
4 2019_04_16_204510_criar_tabela_perfilco2.php
5 2019_04_16_204551_criar_tabela_perfilsolo.php

```

Figura 21: Lista das *migrations* criadas.

Destaca-se a necessidade de se conhecer o equipamento instalado e a maneira como o mesmo exporta seus dados, depois de serem inseridos no sistema. Segue na figura 22 uma parte do código fonte, especificamente entre a linha 12 a 19, apresenta cada coluna da tabela a ser criada pelo Laravel, de acordo com a figura apresentada 18, onde, no diagrama mostra cada tabela com seus respectivos campos.

```

1 <?php
2
3 use Illuminate\Support\Facades\Schema;
4 use Illuminate\Database\Schema\Blueprint;
5 use Illuminate\Database\Migrations\Migration;
6
7 class CriarTabelaNephelometro extends Migration
8 {
9     public function up()
10    {
11        Schema::create('nephelometro', function (Blueprint $table) {
12            $table->increments('id');
13            $table->dateTime('dataNephelometro');
14            $table->string('typeNephelometro');
15            $table->string('coefNephelometro');
16            $table->float('airtempNephelometro');
17            $table->float('celltempNephelometro');
18            $table->float('rhNephelometro');
19            $table->float('pressureNephelometro');
20            $table->timestamps();
21        });
22    }
23    public function down()
24    {
25        Schema::dropIfExists('nephelometro');
26    }
27 }

```

Figura 22: Parte do código de uma *migration*.

O próximo passo foi a criação de modelos, pois cada tabela do banco de dados possui um modelo correspondente, utilizado para interagir com essa tabela, Os modelos permitem consultar dados em suas tabelas, além de inserir novos

registros. Para criar-se um novo modelo, executa-se o comando apresentado na figura 23.

```
$ php artisan make:model nephelometro
```

Figura 23: Comando para criação das *Models*.

FONTE: <https://laravel.com/docs/6.x>

Ao inserir um ID (chave primária do bando de dados) de modelo em uma ação de rota ou controlador, pode-se consultar ou recuperar o modelo que corresponde a esse ID. A ligação do modelo de rota do Laravel fornece uma maneira conveniente de inserir automaticamente as instâncias do modelo diretamente em suas rotas. Por exemplo: em vez de inserir o ID de um usuário, você pode inserir a instância inteira do modelo de Usuário que corresponde ao ID fornecido, como pode ser visto na figura 24.

```
1  <?php
2
3  namespace App;
4
5  use Illuminate\Database\Eloquent\Model;
6
7  class Nephelometro extends Model
8  {
9      protected $table = "nephelometro";
10
11
12     protected $fillable = [
13         'dataNephelometro',
14         'typeNephelometro',
15         'coefNephelometro',
16         'airtempNephelometro',
17         'celltempNephelometro',
18         'rhNephelometro',
19         'pressureNephelometro',
20     ];
21 }
```

Figura 24: Parte do código de uma *Model*.

Com esses passos criados o próximo procedimento de criação foi das *views* (telas). Será apresentada apenas uma delas, mas o código está disponibilizado no GitHub¹ do trabalho para melhor análise.

A figura 25 apresenta o código fonte com uso do Blade (*views*) onde é o mecanismo de modelagem simples, porém poderoso, fornecido com o Laravel.

¹https://github.com/igkuhnen/igor_ufmt

```

1  @extends('layouts.app', ['current' => "nephelometro"])
2
3  @section('content')
4  <div class="container">
5    <div class="card border">
6      <div class="card-body">
7        <h5 class="card-title">Exibindo {{Scats->count()}} Nephelometro de {{Scats->total()}}
8          ( {{Scats->firstItem()}} a {{Scats->lastItem()}} )</h5>
9        @if(count(Scats) > 0)
10       <table class="table table-ordered table-hover">
11         <thead>
12           <tr>
13             <th>Código</th>
14             <th>Data</th>
15             <th>Tipo</th>
16             <th>Coeficiente</th>
17             <th>Temp. do ar</th>
18             <th>Temp Cell </th>
19             <th>RH</th>
20             <th>Pressão</th>
21             <th>Ações</th>
22           </tr>
23         </thead>
24         <tbody>
25           @foreach(Scats as $Scat)
26             <tr>
27               <td>{{ $Scat->id }}</td>
28               <td>{{ $Scat->dataNephelometro }}</td>
29               <td>{{ $Scat->typeNephelometro }}</td>
30               <td>{{ $Scat->coefNephelometro }}</td>
31               <td>{{ $Scat->airtempNephelometro }}</td>
32               <td>{{ $Scat->celltempNephelometro }}</td>
33               <td>{{ $Scat->rhNephelometro }}</td>
34               <td>{{ $Scat->pressureNephelometro }}</td>
35               <td>
36                 <a href="/nephelometro/editar/{{ $Scat->id }}" class="btn btn-sm btn-primary">Editar</a>
37                 <a href="/nephelometro/apagar/{{ $Scat->id }}" class="btn btn-sm btn-danger">Apagar</a>
38               </td>
39             </tr>
40           @endforeach
41         </tbody>
42       </table>
43     @endif
44   </div>
45 </div>
46 <div class="card-footer text-muted">
47   {{Scats->links()}}
48 </div>
49 <div class="card-footer">
50   <a href="/nephelometro/novo" class="btn btn-sm btn-primary" role="button">Novo Nephelometro</a>
51   <a href="/ImpExpNepheView" class="btn btn-sm btn-primary" role="button">Importar/Exportar</a>
52 </div>
53 </div>
54
55
56
57 @endsection

```

Figura 25: Parte do código de uma *View*.

Diferentemente de outros mecanismos de modelagem PHP populares, o Blade não o impede de usar código PHP simples em suas visualizações. De fato, todas as visualizações do Blade são compiladas em código PHP simples e armazenadas em cache até serem modificadas, o que significa que o Blade adiciona uma sobrecarga essencialmente zero ao seu aplicativo. Os arquivos de exibição do blade usam a extensão de arquivo `.blade.php` e geralmente são armazenados no diretório `resources/views`.

O código apresentado acima apresentará conforme apresentado na figura

4.2 APRESENTAÇÃO DO AMBIENTE

Serão demonstrados a seguir a apresentação do sistema, na sua síntese, tais como as telas criadas e sua funcionalidades, área de contato com o usuário final, quem opera todas as funções disponíveis.

Ao iniciar o sistema, apresenta a visão geral do mesmo, com *links* para acesso a informação de cada equipamento, explicando sua funcionalidade e características fundamentais. Pode-se observar isso através da figura 26.



Figura 26: Tela de apresentação do sistema.

Para apresentação dos gráficos utilizou-se o Grafana², uma plataforma de análise de código aberto. O *Link*³ chamado "Gráficos", direciona para o local onde estão hospedados os gráficos gerados no Grafana com a base de dados do PGFA como apresentada na figura 27.

²<<https://grafana.com/>>

³<<http://gpfma.ufmt.br/grafana/d/3-VHYIGWz/monitoramento>>

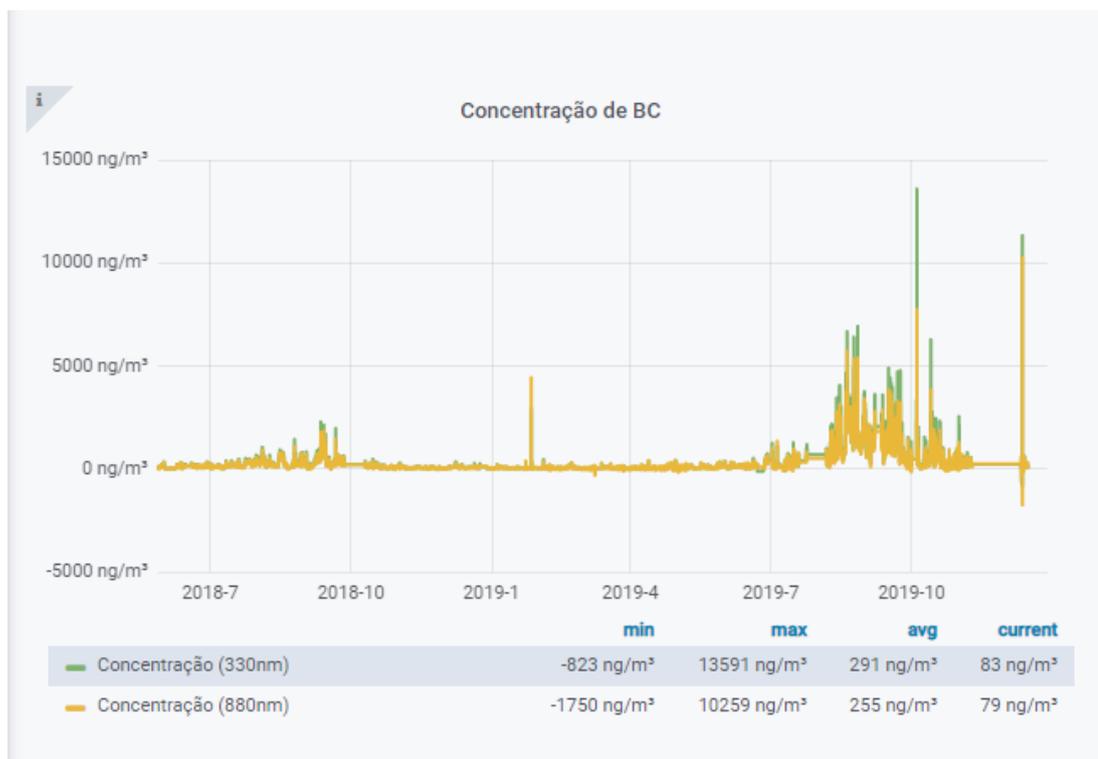


Figura 27: Gráfico de Concentração de *Black Carbon* apresentado pelo Grafana.

A figura 27 apresenta o gráfico do *Grafana* com a concentração de *Black Carbon* que é um dos principais constituintes atmosféricos, pois influencia na qualidade do ar como nas mudanças climáticas (PALACIOS, 2017).

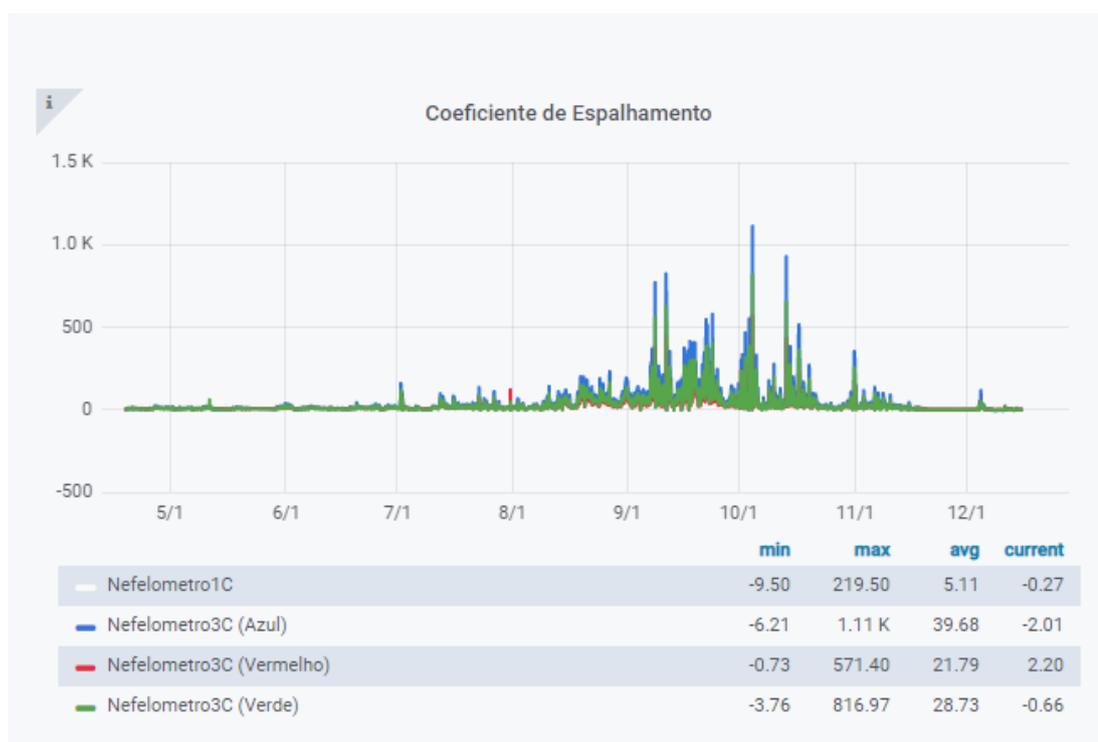


Figura 28: Gráfico de Coeficiente de Espalhamento apresentado pelo Grafana.

A figura 28 apresenta o gráfico do *Grafana* com o coeficiente de espalhamento que é uma medida do Nefelômetro (PALACIOS, 2017).

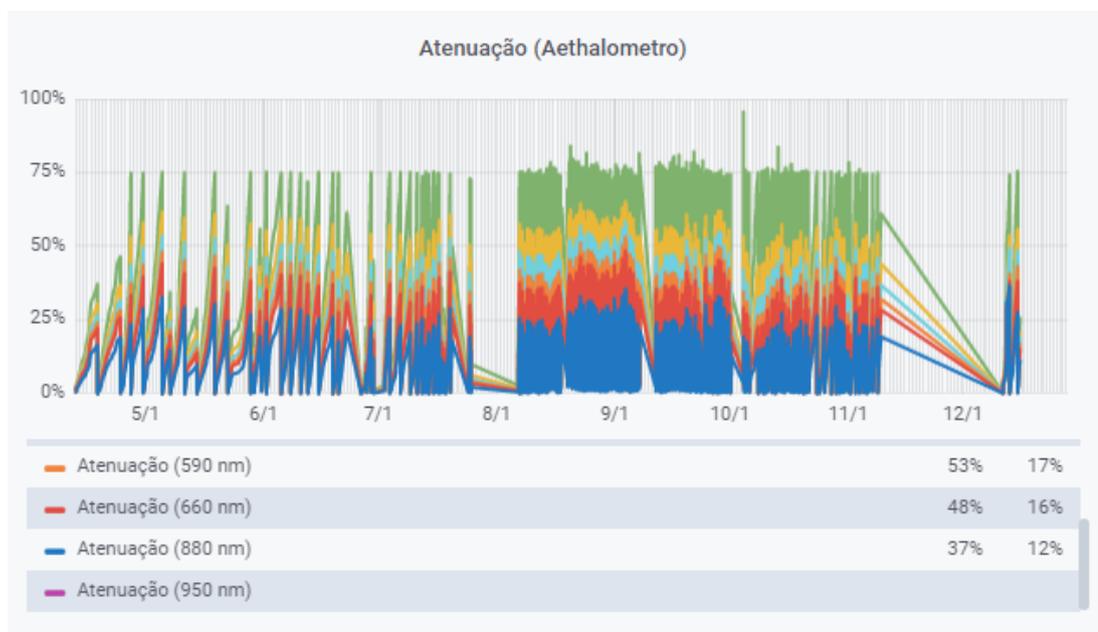


Figura 29: Gráfico de Coeficiente de Espalhamento apresentado pelo Grafana.

A figura 29 apresenta o gráfico do *Grafana* com a medida de atenuação que é coletado pelo Aetalômetro (PALACIOS, 2017).

Ao acessar o link de um dos equipamentos, como citado anteriormente, na figura 30, mostra um exemplo da página criada.



Figura 30: Tela de apresentação e descrição dos equipamentos na BAPP.

Na figura 30 apresenta o instrumento "Aetalômetro" onde o usuário tem acesso a todas as informações específicas deste instrumento, como o princípio de funcionamento e medidas, processo de coleta, resolução temporal das medidas e até mesmo as formas de download dos dados.

4.3 FUNCIONALIDADES DO AMBIENTE

As funcionalidades são a área de administração dos dados coletados pelos sensores, as áreas descritas a seguir ilustra como o agente responsável pelas informações pode processar os dados. Para ter acesso a área é necessário um cadastro para acesso, assim, ao acessar a área de login, como ilustra a figura 31.

Figura 31: Tela de login do sistema para acesso aos dados do sistema.

Esta tela de login, oferece ao usuário a oportunidade de digitar o e-mail e senha para obter acesso, caso tenha permissão, a área administrativa do sistema, onde pode-se visualizar os dados inseridos no sistema para cada equipamento, para que o agente possa escolher onde trabalhar.

Após o acesso serão apresentados os equipamento, como apresenta a figura 32.

Figura 32: Tela de escolha dos dados a serem visualizados para cada equipamento.

A escolha do sensor irá direcionar para a tela contendo todos os dados inclusos na tabela do banco de dados, incluindo as ações para cada linha, onde se pode Editar ou Apagar (figura33).

Código	Data	Tipo	Coefficiente	Temp. do ar	Temp Cell	RH	Pressão	Ações
1	2019-04-29 00:00:00	ert	12	15	12	12	12	Editar Apagar

Figura 33: Tela de exibição de todos os dados do equipamento selecionado no banco de dados.

na figura 33 apresenta várias colunas, onde foram analisadas e implementadas para cada tipo de equipamento, apresentado no Diagrama de Entidade Relacionamento, assim, as linhas conferem os dados deste equipamento.

Tendo necessidade de adicionar um novo dado, deve- acessar o botão de "Novo", e inseri-lo. Assim, apresenta a tela ilustrada na figura 34.

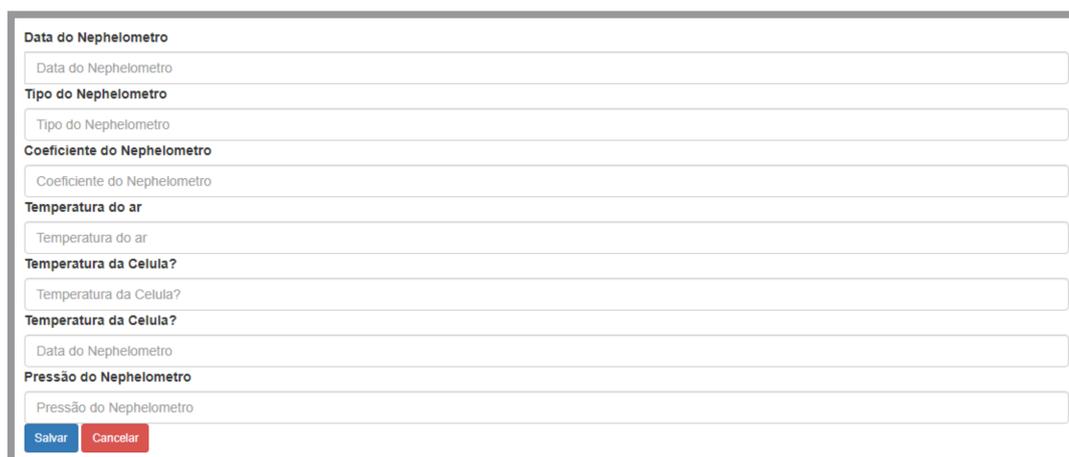


Figura 34: Tela para inserção do novo dados, caso seja necessária alguma manutenção.

Ao abrir a tela do Novo dado, obtêm-se o que tinha como coluna para inserção de um novo dados. Essa tela é usada somente para correção de *outlier* (valor atípico) e/ou calibração do equipamento.

Outra opção que o sistema oferece é de "importação e exportação" dos dados inseridos no sistema. A exportação dos dados está diretamente ligada para que no momento de envio de dados para desenvolvimento de artigos, teses, etc, a integridade dos dados seja mantida. A figura 35 a seguir ilustra essas funções.

No caso da importação é necessária para inserir os dados já obtidos pelo equipamento, mas ainda não consta no sistema. Assim, a cada ampliação de equipamento realizada no sistema, não tenha necessidade de inserir os dados obtidos manualmente.

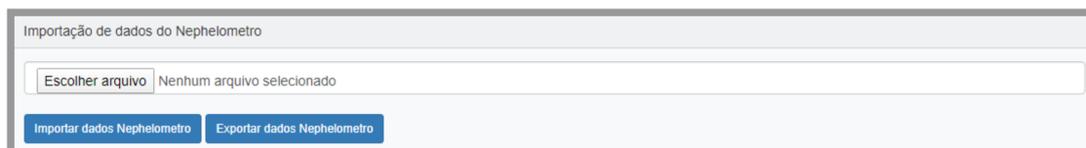


Figura 35: Tela que apresenta o botão para importação dos dados que estavam no equipamento para o banco de dados. A exportação é o *download* de todos os dados que esteja no banco de dados para arquivo de saída.

Capítulo 5

CONCLUSÃO

A análise dos equipamentos e técnicas de processo, desde o planejamento, coleta, armazenamento, análise e produção até a disseminação dos dados ambientais desenvolvidos na BAPP da UFMT, possibilitou a identificação do ciclo de vida atual dos dados e dos pontos críticos que impedem ou dificultam a replicabilidade e integridade dos mesmos.

Por meio de diagramas de fluxo de dados os processos atuais foram compatibilizados a um novo ciclo de vida de dados, englobando as etapas de planejamento, coleta, segurança, descrição, preservação, descoberta, integração, análise e, novamente, o planejamento, possibilitando que o mesmo conjunto de dados gere diferentes informações e conhecimento. Este novo ciclo garante que o conteúdo da informação coletada na estação não se perca ao longo do tempo por questões da aposentadoria ou mudança de carreira dos pesquisadores, preservando ainda a integridade dos dados. Por meio de diagramas de entidade e relacionamento estabeleceram-se os relacionamentos destes dados em uma base de dados e pelos diagramas de atividades descreveram-se o comportamento do aluno/técnico na lógica dos processos.

Uma ferramenta computacional Web foi desenvolvida, em linguagem PHP de uso generalizada, conhecida e de código aberto, em um framework Laravel, de forma a possibilitar a aplicação deste novo ciclo aos processos da BAPP, com descrição dos requisitos para desenvolvimento do sistema, responsáveis pelas regras estabelecidas, desde tipos de equipamentos até recursos a serem respeitados no sistema, formas de consulta, inserção, regras de usuário, compartilhamento dos dados e todos os registros. A simplicidade e transparência do código favorece a manutenção e ampliação do sistema para outros equipamentos da BAPP, bem como para outras bases de pesquisas ambientais.

O sistema gerenciador de banco de dados, MYSQL, foi parametrizado de

forma a garantir elevada robustez e fácil aplicabilidade na replicação dos dados em máquinas distintas.

5.1 TRABALHOS FUTUROS

- Definir estratégias na parte visual, na visualização e análise de extensos gráficos de dados;
- Adaptar o sistema proposto para ambientes de computação de alto desempenho;
- Definir um componente de análise de dados para a arquitetura computacional, com o objetivo de otimizar o processo de compreensão e validação de experimentos e recursos de dados;
- Estender o sistema e a arquitetura computacional para considerar a geração de um grafo do conhecimento, que poderá relacionar dados, imagens, vídeos, publicações, pesquisadores e instituições.

REFERÊNCIAS

BORGMAN, C. L. The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, Wiley Online Library, v. 63, n. 6, p. 1059–1078, 6 2012. ISSN 1532-2890. Disponível em: <<https://doi.org/10.1002/asi.22634>>. Citado na página 7.

BRACKETT, M.; EARLEY, P. S. The dama guide to the data management body of knowledge (dama-dmbok guide). 2009. Citado 3 vezes nas páginas 6, 7 e 10.

BRINEY, K. *Data Management for Researchers: Organize, maintain and share your data for research success*. [S.l.]: Pelagic Publishing Ltd, 2015. Citado 2 vezes nas páginas 7 e 10.

CARDOSO, M. R. A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Rev. bras. epidemiol*, SciELO Public Health, v. 4, n. 3, 2001. Citado na página 4.

COUTINHO, E. R.; SILVA, R. M.; HUBER, F.; DELGADO, A. R. S.; CORREA, B. S. P. M.; VALE, I. G. Desenvolvimento de uma ferramenta de apoio a processamento e monitoramento de dados meteorológicos. *Revista Edu. Tec.*, v. 2, n. 1, 2015. Citado na página 5.

DATE, C. *Introdução a sistemas de bancos de dados*. Campus, 2004. ISBN 9788535212730. Disponível em: <<https://books.google.com.br/books?id=xBeO9LSIK7UC>>. Citado na página 13.

EHLERS, R. S. Análise de séries temporais. *Laboratório de Estatística e Geoinformação*. Universidade Federal do Paraná, 2007. Citado na página 4.

ELMASRI, R.; NAVATHE, S. B.; PINHEIRO, M. G.; CANHETTE, C. C.; MELO, G. C. V.; AMADEU, C. V.; MORAIS, R. de O. *Sistemas de banco de dados*. Pearson Addison Wesley, 2005. Citado na página 13.

FERGUSON, A. R.; NIELSON, J. L.; CRAGIN, M. H.; BANDROWSKI, A. E.; MARTONE, M. E. Big data from small data: data-sharing in the 'long tail' of neuroscience. *Nature neuroscience*, Nature Publishing Group, v. 17, n. 11, p. 1442, 2014. Citado na página 9.

GUJARATI D. N.; PORTER, D. C. *Econometria Básica-5*. [S.l.]: McGraw Hill Brasil, 2011. Citado na página 4.

HEIDORN, P. B. Shedding light on the dark data in the long tail of science. *Library trends*, Johns Hopkins University Press, v. 57, n. 2, p. 280–299, 2008. Citado na página 9.

HOLANDA, B. A. *Absorção da radiação por aerossóis na Amazônia*. Dissertação (Mestrado) — Universidade de São Paulo, 2015. Citado 2 vezes nas páginas 20 e 21.

HOLT, J. *UML for Systems Engineering: watching the wheels*. Institution of Engineering and Technology, 2004. (Computing). Disponível em: <<https://digital-library.theiet.org/content/books/pc/pbpc004e>>. Citado na página 14.

HOWE, B.; COLE, G.; SOUROUSH, E.; KOUTRIS, P.; KEY, A.; KHOUSSAINOVA, N.; BATTLE, L. Database-as-a-service for long-tail science. In: SPRINGER. *International Conference on Scientific and Statistical Database Management*. [S.l.], 2011. p. 480–489. Citado na página 9.

KUHNEN, I. *Análise De Sistemas De Gerenciamento De Banco De Dados Para Armazenamento De Dados Climáticos*. Dissertação (Mestrado) — Mestrado em Física Ambiental, Universidade Federal de Mato Grosso, Instituto de Física, Cuiabá, 2016., 2016. Citado na página 4.

LUCIDCHART. *Lucidchart*. 2020. Disponível em: <<https://www.lucidchart.com/>>. Citado 2 vezes nas páginas 15 e 16.

LYNCH, C. Big data: How do your data grow? *Nature*, Nature Publishing Group, v. 455, n. 7209, p. 28, 2008. Citado na página 9.

MADDALA, G. S.; LAHIRI, K. *Introduction to econometrics*. [S.l.]: Macmillan New York, 1992. v. 2. Citado na página 4.

MEDEIROS, C. B. *Gestão de Dados Científicos - da coleta à preservação*. 2018. Disponível em: <<https://blog.scielo.org/blog/2018/06/22/gestao-de-dados-cientificos-da-coleta-a-preservacao/>>. Citado na página 6.

MICHENER, W. K.; BRUNT, J. W.; HELLY, J. J.; KIRCHNER, T. B.; STAFFORD, S. G. Nongeospatial metadata for the ecological sciences. *Ecological Applications*, Wiley Online Library, v. 7, n. 1, p. 330–342, 1997. Citado 2 vezes nas páginas 8 e 9.

MICHENER, W. K.; PORTER, J.; SERVILLA, M.; VANDERBILT, K. Long term ecological research and information management. *Ecological Informatics*, v. 6, n. 1, p. 13 – 24, 2011. ISSN 1574-9541. Special Issue: 5th Anniversary. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1574954110001159>>. Citado na página 12.

MORESI, E.; PRADO, H.; ALCANTARA, A. Cenários prospectivos, monitoração ambiental e metadados. *Revista de Ciência da Informação*, v. 11, n. 1, 2010. Citado 2 vezes nas páginas 4 e 5.

OLIVEIRA, A. *Plataforma computacional para mineração de dados micrometeorológicos*. 2015. Tese (Doutorado) — Doutorado em Física Ambiental, Universidade Federal de Mato Grosso, Instituto de Física, Cuiabá, 2015.[Links], 2015. Citado na página 4.

PALACIOS, R. d. S. *INTERAÇÃO ENTRE A RADIAÇÃO SOLAR DIRETA E OS AEROSSÓIS ATMOSFÉRICOS NA AMAZONIA*. Tese (Doutorado) — UNIVERSIDADE FEDERAL DE MATO GROSSO, 2017. Citado 4 vezes nas páginas 20, 21, 35 e 36.

PERIS, M. P. M. R. J. Replica control. In: _____. *Encyclopedia of Database Systems*. New York, NY: Springer New York, 2018. p. 3161–3167. ISBN 978-1-4614-8265-9. Disponível em: <https://doi.org/10.1007/978-1-4614-8265-9_310>. Citado na página 13.

PETERSON, A. T.; SOBERÓN, J.; KRISHTALKA, L. A global perspective on decadal challenges and priorities in biodiversity informatics. *BMC ecology*, BioMed Central, v. 15, n. 1, p. 15, 2015. Citado na página 8.

REZENDE, R. *Conceitos Fundamentais de Banco de Dados*. 2006. Disponível em: <<http://www.devmedia.com.br/conceitos-fundamentais-de-banco-de-dados/1649#ixzz3rmMDdwXHX>>. Citado na página 13.

RÜEGG, J.; GRIES, C.; BOND-LAMBERTY, B.; BOWEN, G. J.; FELZER, B. S.; MCINTYRE, N. E.; SORANNO, P. A.; VANDERBILT, K. L.; WETHERS, K. C. Completing the data life cycle: using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment*, Wiley Online Library, v. 12, n. 1, p. 24–30, 2014. Citado na página 10.

SPECIFICATION, O. A. Omg unified modeling language (omg uml), superstructure, v2. 1.2. *Object Management Group*, v. 70, 2007. Citado na página 14.

STRASSER, C.; COOK, R.; MICHENER, W.; BUDDEN, A. Primer on data management: what you always wanted to know. 2012. Citado 2 vezes nas páginas 7 e 11.

VENTURA, P. *Ateomomento*. 2020. Disponível em: <<https://www.ateomomento.com.br/>>. Citado na página 15.

VIANA, A. L. d. S. *MySQL: Replicação de Dados*. 2011. Disponível em: <<https://www.devmedia.com.br/mysql-replicacao-de-dados/22923>>. Citado na página 14.

VINES, T. H.; ALBERT, A. Y.; ANDREW, R. L.; DÉBARRE, F.; BOCK, D. G.; FRANKLIN, M. T.; GILBERT, K. J.; MOORE, J.-S.; RENAUT, S.; RENNISON, D. J. The availability of research data declines rapidly with article age. *Current Biology*, v. 24, n. 1, p. 94 – 97, 2014. ISSN 0960-9822. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960982213014000>>. Citado na página 10.

WALLIS, J. C.; ROLANDO, E.; BORGMAN, C. L. If we share data, will anyone use them? data sharing and reuse in the long tail of science and technology. *PloS one*, Public Library of Science, v. 8, n. 7, p. e67332, 2013. Citado na página 9.