

**Programa de Monitoramento da Biodiversidade Aquática da
Área Ambiental I – Porção Capixaba do Rio Doce e Região
Marinha e Costeira Adjacente**

PLANO DE GESTÃO DE DADOS (PGD)

PGD RRDM/JUN19

Coordenação Geral

Adalto Bianchini

Alex Cardoso Bastos

Edmilson Costa Teixeira

Eustáquio Vinícius de Castro

Jorge Abdala Dergam dos Santos

Vitória,

Junho de 2019

COORDENAÇÕES

COORDENAÇÕES POR ANEXO

Anexo 1

Adalto Bianchini (FURG)

Anexo 3

Fabian Sá (UFES)

Gilberto Fonseca Barroso (UFES)

Subprojetos

Alessandra Delazari Barroso (FAESA)

Alex Cardoso Bastos (UFES)

Ana Cristina Teixeira Bonecker (UFRJ)

Anderson Geyson Alves de Araújo (UFES)

Björn Gücker (UFSJ)

Camilo Dias Júnior (UFES)

Daniel Rigo (UFES)

Edmilson Costa Teixeira (UFES)

Eneida Maria Eskinazi Sant'Anna (UFOP)

Gilberto Amado Filho (IPJB)

Iola Gonçalves Boechat (UFSJ)

Leila Lourdes Longo (UFRB)

Luís Fernando Loureiro (UFES)

Marco Aurélio Caiado (UFES)

Renato David Ghisolfi (UFES)

Renato Rodrigues Neto (UFES)

Rodrigo Leão de Moura (UFRJ)

Valéria da Silva Quaresma (UFES)

Valéria de Oliveira Fernandes (UFES)

Vanya Marcia Duarte Pasa (UFMG)

Anexo 4

Jacqueline Albino (UFES)

Subprojetos

Karla Costa (UFES)

Maria Tereza Carneiro (UFES)

Anexo 5

Diolina Moura Silva (UFES)

Mônica Tognella (UFES)

Anexo 6

Agnaldo Silva Martins (UFES)

Subprojetos

Ana Paula Cazerta Farro (UFES)

Leandro Bugoni (FURG)

Sarah Vargas (UFES)

Anexo 7

Maurício Hostim (UFES)

Jorge Dergam (UFV)

Subprojetos

Carlos W. Hackradt (UFSB)

Fabiana Felix Hackradt (UFSB)

Jean-Christophe Joyeux (UFES)

Luis Fernando Duboc (UFV)

Anexo 8

Heitor Evangelista (UERJ)

Coordenação Núcleo de Atuação Integrada em Rede

Edmilson Costa Teixeira (UFES)

Karla Libardi Gallina (UFES)

Andressa Christiane Pereira (UFES)

Anna Paula Lage Ribeiro (UFES)

Caroline De Marchi Pignaton (UFES)

Paulo Eduardo Marques (UFES)

Coordenação Técnica

Alex Cardoso Bastos (UFES)

Laura Silveira Vieira Salles (UFES)

Tarcila Franco Menandro (UFES)

Coordenação Escritório de Projetos

Eustáquio Vinicius Ribeiro de Castro (UFES)

Patrícia Bourguignon Soares (UFES)

Armando Biondo Filho (UFES)

Paulo Roberto Filgueiras (UFES)

Valdemar Lacerda Junior (UFES)

Walter Luiz Alda Junior (UFES)

SUMÁRIO

1	SUMÁRIO EXECUTIVO	10
1.1	CONTEXTO DE REFERÊNCIA.....	10
1.2	FUNDAÇÃO RENOVA.....	10
1.3	REDE RIO DOCE MAR (RRDM).....	11
1.4	PROGRAMA DE MONITORAMENTO DA BIODIVERSIDADE AQUÁTICA (PMBA).....	12
1.4.1	Objetivos	13
1.4.2	Resultados Esperados	13
1.4.3	Período de Execução	14
1.4.4	Projetos e Anexos do Programa	14
1.4.5	Coordenação e Equipe dos Anexos e Subprojetos	15
1.5	UNIDADE DE GESTÃO DE DADOS (UGD).....	17
1.6	UNIDADE DE ATUAÇÃO EM REDE (UAR).....	18
1.7	PLANO DE GESTÃO DE DADOS (PGD).....	19
2	ARQUITETURA DE DADOS	20
2.1	CAMADAS DE GESTÃO DE DADOS.....	21
2.1.1	Origem de Dados	22
2.1.2	Armazém de Dados	23
2.1.3	Repositório de Dados	24
2.1.4	Banco de Dados	27
2.1.5	Análise de Dados	30
2.2	REQUISITOS DE SOFTWARE.....	30
2.2.1	Armazém de Dados	30
2.2.2	Repositório de Dados	32
2.2.3	Banco de Dados	33
2.3	REQUISITOS DE HARDWARE.....	33

2.3.1	Armazém de Dados	33
2.3.2	Repositório de Dados	35
2.3.3	Banco de Dados	36
3	GESTÃO DE DADOS	37
3.1	COLETA DE DADOS	37
3.1.1	Quais dados serão coletados ou criados?	37
3.1.2	Como os dados serão coletados ou criados?	44
3.2	DOCUMENTAÇÃO E METADADOS	51
3.2.1	Quando se deve iniciar a documentação dos dados?	51
3.2.2	Quais as ferramentas e padrões de documentação de metadados serão utilizados? ..	52
3.2.3	Que documentação e metadados acompanharão os dados?	52
3.2.4	Como identificar unicamente cada conjunto de dados?	56
3.3	ÉTICA E DIREITOS AUTORAIS	57
3.3.1	Como serão administradas as questões éticas?	57
3.3.2	Como serão gerenciados os direitos autorais?	57
3.4	SELEÇÃO E PRESERVAÇÃO	60
3.4.1	Quais dados são de valor a longo prazo e devem ser mantidos, compartilhados e/ou preservados?	60
3.4.2	Qual é o plano de preservação a longo prazo do conjunto de dados (dataset)?	61
3.5	ARMAZENAMENTO E BACKUP	62
3.5.1	Como os dados serão armazenados durante a pesquisa?	62
3.5.2	Como serão realizados os backups de dados durante a pesquisa?	63
3.5.3	Como serão gerenciados os acessos e a segurança?	64
3.6	COMPARTILHAMENTO DE DADOS	65
3.6.1	Como os dados serão compartilhados?	66
3.6.2	Existem restrições quanto ao compartilhamento de dados?	68
3.6.3	Como dar o crédito adequado ao produtor dos dados?	69

3.7	RESPONSABILIDADES E RECURSOS	70
3.7.1	Quem será responsável pelo gerenciamento de dados?	70
3.7.2	Quais recursos serão necessários para implementar o plano de gestão de dados?	72
4	INTEGRAÇÃO UGD E UAR	74
4.1	SISTEMA WEB DE AUXÍLIO À ATUAÇÃO EM REDE/RRDM (SW-A2R/RRDM).....	75
4.1.1	Informações Abertas à Sociedade	75
4.1.2	Funcionalidades Restritas aos Colaboradores	76
4.2	INTEGRAÇÃO DE AÇÕES E PRODUTOS.....	77
5	CUSTOS DE IMPLEMENTAÇÃO	80
5.1	INFRAESTRUTURA.....	80
5.1.1	Custos de Software.....	80
5.1.2	Custos de Hardware.....	80
5.2	RECURSOS HUMANOS	82
5.2.1	Equipe de Implantação	82
5.2.2	Equipe de Suporte	83
5.2.3	Treinamentos	83
6	CRONOGRAMA DE IMPLEMENTAÇÃO.....	83
6.1	ANO 1 (ATUAL).....	86
6.2	ANO 2	87
6.3	ANOS 3, 4 E 5.....	87
7	CONSIDERAÇÕES FINAIS	87
8	ANEXOS	89

LISTA DE FIGURAS

Figura 1: Camadas de Gestão de Dados previstas na infraestrutura do PMBA.....	21
Figura 2: Inclusão do PMBA como um nó da rede internacional.	26
Figura 3: Arquitetura da camada de Banco de Dados, subcamadas e ferramentas.....	29
Figura 4: Estrutura interna de pastas (primeiro nível) do Anexo 04 (Praias).	48
Figura 5: Estrutura interna de pastas (segundo nível) do Anexo 04 (Praias).	48
Figura 6: Estrutura interna de pastas (terceiro nível) do Anexo 04 (Praias).	48
Figura 7: Estrutura interna de pastas (quarto nível) do Anexo 04 (Praias).....	49
Figura 8: Tela do Metacat para cadastro de informações gerais do data set.	53
Figura 9: Tela do Metacat para cadastro das pessoas associadas ao dataset.	53
Figura 10: Tela do Metacat para cadastro da cobertura temporal do dataset.	54
Figura 11: Tela do Metacat para cadastro da cobertura geográfica do dataset.....	55
Figura 12: Tela do Metacat para cadastro da cobertura taxonômica do dataset.	55
Figura 13: Tela do Metacat Data Catalog para cadastro das metodologias utilizadas.....	56
Figura 14: Tela principal do Sistema Web da Rede Rio Doce Mar.	75
Figura 15: Tela principal da área restrita aos colaboradores da RRDM.....	77
Figura 16: Fluxograma integrativo de ações e produtos entre as Unidades UGD e UAR.....	79

LISTA DE TABELAS

Tabela 1: Custos dos profissionais alocados na equipe de implantação	82
Tabela 2: Custos dos profissionais alocados na equipe de suporte.....	83

LISTA DE QUADROS

Quadro 1: Lista de instituições integrantes da RRDM	11
Quadro 2: Levantamento de volumetria de dados para cada um dos Anexos/Projetos.....	34
Quadro 3: Levantamento das ferramentas para geração de dados pelos Anexos/Projetos.	39
Quadro 4: Levantamento dos formatos de dados a serem gerados pelos Anexos/Projetos	40
Quadro 5: Usuários ativados no Dropbox e contatos dos respectivos responsáveis.....	45
Quadro 6: Grupos de usuários criados no Dropbox e usuários associados	46
Quadro 7: Pastas de equipe e permissões definidas no Armazém de Dados (Dropbox).....	47
Quadro 8: Áreas de armazenamento e de backup estimadas para as camadas.	62
Quadro 9: Cronograma de implementação do Plano de Gestão de Dados.....	84

1 SUMÁRIO EXECUTIVO

O objetivo deste documento é apresentar o Plano de Gestão de Dados (PGD) para o Programa de Monitoramento da Biodiversidade Aquática (PMBA) em execução pela Rede Rio Doce Mar (RRDM).

Nesta seção são apresentadas maiores informações sobre o Programa de Monitoramento da Biodiversidade Aquática (PMBA), incluindo um breve contexto de referência, objetivos, resultados esperados, prazos e projetos associados. Também são apresentadas a Fundação Renova e a Rede Rio Doce Mar (RRDM). Por fim, é abordada a importância da Unidade de Gestão de Dados (UGD) e do Plano de Gestão de Dados (PGD) para o Programa, e são descritos cada um dos tópicos da estrutura do PGD.

1.1 CONTEXTO DE REFERÊNCIA

Em 5 de novembro de 2015 houve o rompimento da barragem de rejeitos de minério de ferro de Fundão, na localidade de Mariana (MG), deslocando volume de contaminantes ao longo de quase 700 km até à foz do rio Doce na localidade de Regência (ES) e na plataforma continental adjacente.

Em março de 2016, por meio de um Termo de Transação e Ajustamento de Conduta (TTAC), assinado entre Samarco, seus acionistas, os governos federal e dos estados de Minas Gerais e do Espírito Santo, foi constituída a Fundação Renova com o propósito de gerir e executar os programas e ações para recuperar, remediar e reparar os impactos gerados a partir do rompimento da Barragem de Fundão.

O Programa de Monitoramento da Biodiversidade Aquática (PMBA) da Área Ambiental I faz parte do acordo de cooperação firmado entre a Fundação Espírito-Santense de Tecnologia (FEST) e a Fundação Renova e dará subsídios para a tomada de decisão das ações de conservação da biodiversidade da região impactada pelo rompimento da Barragem de Fundão, em Mariana (MG). A Rede Rio Doce Mar (RRDM) é a responsável pela execução do PMBA na porção espírito-santense da bacia do rio Doce e regiões costeira adjacente e marinha.

Para maiores informações sobre o Acordo que deu origem ao PMBA, favor consultar os anexos 03 a 10 (Termo de Referência do PMBA), disponibilizado na seção ANEXOS.

1.2 FUNDAÇÃO RENOVA

A Fundação Renova é a entidade responsável pela mobilização para a reparação dos danos causados pelo rompimento da barragem de Fundão, em Mariana (MG). Trata-se de uma organização sem fins lucrativos, resultado de um compromisso jurídico chamado Termo de Transação e Ajustamento de Conduta (TTAC).

Para maiores informações sobre a Fundação Renova, favor consultar o anexo 01 (Estatuto da Fundação Renova), disponibilizado na seção ANEXOS.

1.3 REDE RIO DOCE MAR (RRDM)

A Rede Rio Doce Mar nasceu do comprometimento colaborativo de pesquisadores acadêmicos com a preservação da biodiversidade aquática da área afetada pelo rompimento da barragem de Fundão, em Marina (MG).

Ao todo, são pesquisadores de 25 instituições de ensino estaduais e federais unindo forças para desenvolver um trabalho de coleta e análise de dados contínuo e transparente. A lista de instituições integrantes da RRDM é apresentada no Quadro 1. Para maiores informações sobre a RRDM, favor consultar o anexo 02 (Apresentação da RRDM), disponibilizado na seção ANEXOS.

Quadro 1: Lista de instituições integrantes da RRDM

Sigla	Instituição
FAESA	FAESA Centro Universitário
FURG	Fundação Universidade Federal do Rio Grande
IFES	Instituto Federal do Espírito Santo
JBRJ	Jardim Botânico do Rio de Janeiro
UENF	Universidade Estadual do Norte Fluminense
UERJ	Universidade do Estado do Rio de Janeiro
UFAL	Universidade Federal de Alagoas
UFAM	Universidade Federal do Amazonas
UFES	Universidade Federal do Espírito Santo
UFF	Universidade Federal Fluminense
UFJF	Universidade Federal de Juiz de Fora
UFMG	Universidade Federal de Minas Gerais
UFOP	Universidade Federal de Ouro Preto
UFPE	Universidade Federal de Pernambuco
UFPEL	Universidade Federal de Pelotas
UFPR	Universidade Federal do Paraná
UFRB	Universidade Federal do Recôncavo da Bahia
UFRJ	Universidade Federal do Rio de Janeiro
UFSB	Universidade Federal do Sul da Bahia
UFSJ	Universidade Federal de São João del-Rei
UFSM	Universidade Federal de Santa Maria
UFV	Universidade Federal de Viçosa
UNICAMP	Universidade Estadual de Campinas
UNIVALI	Universidade do Vale do Itajaí
USP	Universidade de São Paulo

Fonte: autoria própria.

1.4 PROGRAMA DE MONITORAMENTO DA BIODIVERSIDADE AQUÁTICA (PMBA)

O Programa de Monitoramento da Biodiversidade Aquática (PMBA) da Área Ambiental 1 engloba o monitoramento da qualidade da água e estudos ecotoxicológicos para verificar a ocorrência de contaminação na biodiversidade aquática na Área Ambiental 1, para subsidiar:

- A tomada de decisões referentes à captura e consumo.
- O processo de avaliação do estado de conservação da biodiversidade aquática.
- A adoção de medidas de recuperação e conservação da fauna e do ambiente aquático.

Para isso, o estudo irá analisar dados anteriores ao rompimento (dados pretéritos) e monitorar a fauna e a flora da porção capixaba do rio Doce e ambientes marinhos e costeiros impactados. Será possível identificar o impacto agudo e crônico sobre as espécies e a cadeia alimentar, além de avaliar o habitat de fundo marinho, a qualidade da água e a ecotoxicidade, a fim de indicar eventuais medidas reparatórias.

No primeiro ano de projeto, serão coletadas amostras de água, sedimentos, animais e vegetais. Para realização dos monitoramentos, estarão envolvidos cerca de 550 profissionais, entre eles colaboradores acadêmicos. É priorizada a contratação de mão de obra local, como pescadores e catadores de caranguejo.

Ao todo serão monitorados pontos de toda a porção capixaba do rio Doce e da região que vai do entorno de sua foz a Guarapari (ES), ao sul, e até Abrolhos (BA), ao norte.

As coletas incluem água, sedimento, organismos aquáticos em geral, incluindo plâncton, algas, bentos, peixes, análises em tartarugas, golfinhos, observação de baleias, aves, até imageamento do fundo marinho.

Para o estudo serão utilizados drones, aeronaves e embarcações de pequeno, médio e grande portes, além de sensores de diversos tipos, imagens de satélites e boias automatizadas, equipadas com instrumentos específicos para esse tipo de monitoramento. Os laboratórios das universidades estão aptos para fazer a análise das amostras.

Os resultados dos monitoramentos serão entregues trimestralmente à Fundação Renova, ao CIF (Comitê Interfederativo) e à CT-Bio (Câmara Técnica de Conservação e Biodiversidade), e a cada semestre será realizado um workshop para apresentação dos resultados, com a participação dos órgãos ambientais.

Para maiores informações sobre o PMBA, favor consultar o anexo 11 (Plano de Trabalho Geral do PMBA), disponibilizado na seção **Erro! Fonte de referência não encontrada.**

1.4.1 Objetivos

O PMBA tem como objetivo geral realizar ações de monitoramento visando verificar a evolução do impacto do rejeito nos ecossistemas e na biodiversidade associada.

Como objetivos específicos, podemos citar:

- Analisar dados pretéritos de monitoramento, do período novembro 2015 até julho de 2018.
- Monitorar a fauna da foz do Rio Doce e ambientes estuarinos e marinhos impactados.
- Identificar e caracterizar o impacto agudo e crônico sobre as espécies e cadeia trófica dos ambientes dulcícolas, estuarino e marinho.
- Avaliar o habitat de fundo marinho, incluindo algas calcárias, rodólitos e corais, nas áreas estuarinas, marinhas e da foz do rio atingidas pelo material oriundo do rompimento da barragem.
- Avaliar a qualidade da água e a ecotoxicidade sobre os organismos aquáticos, estuarinos, marinhos e dulcícolas.
- Subsidiar as ações de conservação da biodiversidade no âmbito da região impactada.

1.4.2 Resultados Esperados

Dentre os resultados esperados, destacam-se os seguintes:

- Definição de bioindicadores da qualidade do ecossistema e do impacto do desastre.
- Entendimento dos impactos agudos e crônicos do desastre nos diferentes ecossistemas, a partir de uma análise integrada usando dados de antes do desastre, quando existirem, e dados obtidos após o rompimento da barragem.
- Entendimento do impacto do desastre na estrutura de flora e fauna nos ambientes estudados, usando dados pretéritos.
- Mapeamento dos habitats marinhos e avaliação temporal dos impactos no mesmo;
- Compreensão do impacto do rejeito na biodiversidade marinha, estuarina e dulcícola.
- Simulação do efeito da pluma de rejeitos no comportamento biogeoquímico e avaliação de sua influência na biodiversidade.
- Diagnóstico do comprometimento da estrutura e função dos ecossistemas aquáticos decorrente do rompimento da barragem.
- Proposição de metas a serem alcançadas para melhorar a qualidade dos ecossistemas.

Proposição de ações de mitigação do impacto ambiental e visando melhor manejo para conservação da biodiversidade.

1.4.3 Período de Execução

Conforme definido nos termos da cláusula 165 do Termo de Transação e de Ajustamento de Conduta (TTAC), o Programa de Monitoramento deverá ser implementado e executado em um período de 5 anos. Após este período, deverá ser avaliada a necessidade de continuidade do Programa vis-à-vis os impactos detectados no período.

A vigência do atual Acordo de Cooperação Fundação Renova – FEST é de 16 meses, referente aos primeiros 12 meses do PMBA; ou seja, 2 meses de mobilização, 12 meses de monitoramento e 2 meses para prestação de contas desta fase.

1.4.4 Projetos e Anexos do Programa

Para alcançar o objetivo proposto, o PMBA foi dividido em 8 grandes temáticas de acordo com o tipo de estudo, ambiente ou grupos de organismos. Cada uma delas é subdividida por expedições de coletas e estudos específicos voltados ao seu campo de atuação.

Estas temáticas, apresentadas abaixo, são referenciadas como Projetos ou Anexos do Programa:

Anexo 1 – Ecotoxicologia: Monitoramento ecotoxicológico dos impactos causados pela lama oriunda do rompimento da barragem de Fundão, Mariana (MG), em regiões estuarinas, marinhas e dulcícola.

Anexo 3 – Marinho-Estuarino: Estudo e monitoramento ambiental da área estuarina e marinha (Bentos; Fitoplâncton; Fundos Recifais; Hidrogeoquímica; Ictioplâncton; Mapeamento de Habitats; Modelagem; Sedimentologia e Zooplâncton).

Anexo 3 – Dulcícola-Integração: Estudo e monitoramento ambiental da área dulcícola no ES (Bacia Integração; Bacia Integração-Sedimentos; Bacia Integração-Gestão Integrada; Bacia Integração-Modelagem; Dulcícola-Elementar Isótopos; Dulcícola-Fitoplâncton; Dulcícola-Macrófita; Dulcícola-Perifíton e Dulcícola-Zooplâncton).

Anexo 4 – Praias: Monitoramento de potenciais impactos do rejeito de minério de ferro na praia e antepraia adjacentes da desembocadura do Rio Doce (Bentos Praias; Morfodinâmica e Geoquímica).

Anexo 5 – Restinga e Manguezal: Alterações ecológicas na dinâmica dos manguezais e vegetação de restinga sob influência dos sedimentos provenientes do Rio Doce.

Anexo 6 – Megafauna: Monitoramento de mamíferos marinhos, tartarugas e aves marinhas associados à foz do rio Doce, plataforma continental e áreas protegidas adjacentes (Sobrevôo e Acústica; Cetáceos; Aves Marinhas e Tartarugas Marinhas).

Anexo 7 – Ictiofauna Marinha e Ictiofauna Dulcícola: Estudo e monitoramento da ictiofauna marinha (Peixes Recifais), estuarina (Ictiofauna Estuarina) e dulcícola (Dulcícola Genética e Ecologia).

Anexo 8 – Sedimentação em Abrolhos: Monitoramento da sedimentação no Parque Nacional Marinho dos Abrolhos e regiões relacionadas.

Para maiores informações sobre os Anexos/Projetos do PMBA, favor consultar os anexos 12 a 19 (Planos de Trabalho dos Anexos do PMBA), disponibilizado na seção **Erro! Fonte de referência não encontrada.** Nos Planos de Trabalho são apresentados os Objetivos, Metas e Justificativas, Equipe Técnica, Produtos (Dados Brutos e Análise de Dados) e Metodologia.

1.4.5 Coordenação e Equipe dos Anexos e Subprojetos

O presente projeto é composto por duas grandes partes, que são executadas concomitantemente:

- Atividades-suporte ao Programa de Monitoramento.

- Programa de Monitoramento para ambientes estuarinos, marinhos e dulcícola, conforme discriminado anteriormente.

As atividades de suporte ao PMBA são realizadas pelas áreas a seguir, com os respectivos coordenadores.

- **Coordenação Técnica (CTEC):** Área responsável pela coordenação geral dos Projetos/Anexos do PMBA.

Coordenação: Alex Cardoso Bastos

- **Escritório de Projetos (ESCPRO):** Área responsável pela Gestão de Projetos, Gestão de Aquisições e Gestão de Recursos Humanos do PMBA.

Coordenação: Eustáquio Vinicius Ribeiro de Castro

- **Núcleo de Atuação Integrada em Rede (NAIR):** Área responsável pela Gestão de Dados, Comunicação, Relacionamento com Parceiros e Análise Integrada do PMBA.

Coordenação: Edmilson Costa Teixeira

Abaixo, são apresentados os coordenadores dos Projetos/Anexos do PMBA.

- **Anexo 1:**

Coordenação Geral: Adalto Bianchini (FURG)

- **Anexo 3:**

Coordenação Geral (Marinho-Estuarino): Fabian Sá (UFES)

Coordenação Geral (Dulcícola-Integração): Gilberto Fonseca Barroso (UFES)

Coordenação de Subprojetos:

- Alessandra Delazari Barroso (FAESA)
- Alex Cardoso Bastos (UFES)
- Ana Cristina Teixeira Bonecker (UFRJ)

- Anderson Geyson Alves de Araújo (UFES)
- Björn Gücker (UFSJ)
- Camilo Dias Júnior (UFES)
- Daniel Rigo (UFES)
- Edmilson Costa Teixeira (UFES)
- Eneida Maria Eskinazi Sant'Anna (UFOP)
- Gilberto Amado Filho (IPJB)
- Iola Gonçalves Boechat (UFSJ)
- Leila Lourdes Longo (UFRB)
- Luís Fernando Loureiro (UFES)
- Marco Aurélio Caiado (UFES)
- Renato David Ghisolfi (UFES)
- Renato Rodrigues Neto (UFES)
- Rodrigo Leão de Moura (UFRJ)
- Valéria da Silva Quaresma (UFES)
- Valéria de Oliveira Fernandes (UFES)
- Vanya Marcia Duarte Pasa (UFMG)

- Anexo 4:

Coordenação Geral: Jacqueline Albino (UFES)

Coordenação de Subprojetos:

- Karla Costa (UFES)
- Maria Tereza Carneiro (UFES)

- Anexo 5:

Coordenação Geral (Restinga): Diolina Moura Silva (UFES)

Coordenação Geral (Manguezal): Mônica Tognella (UFES)

- Anexo 6:

Coordenação Geral: Agnaldo Silva Martins (UFES)

Coordenação de Subprojetos:

- Ana Paula Cazerta Farro (UFES)
- Leandro Bugoni (FURG)
- Sarah Vargas (UFES)

- Anexo 7:

Coordenação Geral (Ictiofauna Marinha): Maurício Hostim (UFES)

Coordenação Geral (Ictiofauna Dulcícola): Jorge Dergam (UFV)

Coordenação de Subprojetos:

- Carlos W. Hackradt (UFSB)
- Fabiana Felix Hackradt (UFSB)
- Jean-Christophe Joyeux (UFES)
- Luis Fernando Duboc (UFES)

- Anexo 8:

Coordenação Geral: Heitor Evangelista (UERJ)

Para visualização dos demais integrantes das equipes de cada Anexo/Projeto, favor consultar o anexo 20 (Equipes do PMBA), disponibilizado na seção ANEXOS. No documento são apresentadas as equipes técnicas, equipes de laboratório e equipes de campo de cada Anexo/Projeto.

1.5 UNIDADE DE GESTÃO DE DADOS (UGD)

Programas de monitoramento da biodiversidade, como o PMBA, produzem uma grande quantidade de dados que se adequadamente geridos (formatados, padronizados e organizados) são capazes de apoiar a elaboração e a implementação de medidas para a recuperação e conservação dos ecossistemas.

Dados de inventários biológicos podem ser muito variáveis dependendo das espécies analisadas, parâmetros monitorados, abrangência espacial e ecossistemas impactados, trazendo grande complexidade para a sua gestão. Por isso, a RRDM conta com uma Unidade de Gestão de Dados (UGD), compondo o Núcleo de Atuação Integrada em Rede (NAIR) do projeto, e tendo como objetivo organizar, integrar, disponibilizar e capacitar pessoas para gerenciar os dados e metadados, de forma que os dados sejam mantidos com segurança, e estejam disponíveis para os tomadores de decisão e pesquisadores ao longo do tempo.

Abaixo, algumas das principais atividades da equipe de Gestão de Dados:

- Especificação de padrões e formatos para dados e metadados.

- Definição de conjunto mínimo de informações sobre dados (metadados).
- Curadoria e organização de conteúdo.
- Implementação de políticas de segurança (acesso, compartilhamento e reuso).
- Oferta de acesso contínuo ao longo do ciclo de vida previsto para os dados.
- Armazenamento, arquivamento e preservação dos dados ao longo de seu ciclo de vida.
- Definição de papéis e responsabilidades a todos os envolvidos com os dados.
- Planejamento de eventuais transições de plataformas, sistemas e equipe envolvidas.
- Definição de infraestrutura e arquitetura de hardware e software necessária.
- Gestão de orçamento (custos) de pessoal, treinamentos e infraestrutura.
- Gestão de Citações, Direitos Autorais e Responsabilidades sobre os dados.
- Treinamentos e capacitações para apoiar o gerenciamento de dados e metadados.

Todas as atividades acima deverão estar descritas no Plano de Gestão Dados (PGD) do Programa, principal produto a ser gerado no primeiro ano do PMBA pela UGD/NAIR.

A RRDM, com o apoio da UGD/NAIR, disponibilizará toda a infraestrutura necessária para o gerenciamento dos dados, inicialmente com a utilização da ferramenta Dropbox para depósito das informações de monitoramento utilizando planilhas de dados formatadas e previamente padronizadas, e futuramente com a publicação dos dados e metadados em repositórios de dados e bancos de dados para permitir consultas, integrações e análises integradas de impacto via ferramentas analíticas.

Para maiores informações sobre a UGD/NAIR, favor consultar o anexo 21 (Planos de Trabalho da UGD), disponibilizado na seção **Erro! Fonte de referência não encontrada.** No Plano de Trabalho são apresentados os Objetivos, Metas e Justificativas, Equipe Técnica, Produtos (Dados Brutos e Análise de Dados) e Metodologia.

1.6 UNIDADE DE ATUAÇÃO EM REDE (UAR)

A UAR, também sob a coordenação do NAIR, é responsável pela aplicação e desenvolvimento de dispositivos e estratégias voltadas para a atuação integrada em rede no âmbito de toda a Estrutura Organizacional de Execução do PMBA. A unidade tem atuado muito próximo da UGD, principalmente em atividades que requerem a utilização de sistemas informacionais.

Dentre suas principais funções, a UAIR disponibilizará um dispositivo web de auxílio à atuação integrada em rede, para emprego no âmbito do PMBA. Na busca do desenvolvimento da versão do dispositivo web de auxílio a atuação integrada em rede (SWAR-RRDM), foi elaborada a especificação de funcionalidades para compartilhar e organizar informações de modo institucional. As funcionalidades

iniciais tiveram como base trabalhos já realizados pelo Laboratório de Gestão em Recursos Hídricos e Desenvolvimento Regional da UFES (LabGest) em dispositivos de aproximação entre grupos de trabalho, além de referências na literatura e análise de pontos de integração com as ferramentas de gestão de dados a serem utilizadas no PMBA. De posse dessas informações foi possível a construção dos requisitos de uma versão inicial do SWAR-RRDM que será utilizado de forma integrada aos outros sistemas informacionais a serem disponibilizados pela Unidade de Gestão de Dados (UGD).

Os pontos de integração entre os sistemas serão apresentados e detalhados ao longo deste documento.

1.7 PLANO DE GESTÃO DE DADOS (PGD)

Um plano de gerenciamento de dados é um documento formal que descreve o que será realizado com os dados durante e após um projeto de pesquisa. A maioria dos pesquisadores coleta dados com objetivos definidos, mas muitas vezes não pensam na documentação e gestão dos dados. Muitos problemas de gerenciamento de dados podem ser resolvidos facilmente ou evitados completamente pelo planejamento antecipado. Com o processo e estrutura corretos, não demora muito e pode render-se enormemente a longo prazo.

O Plano de Gestão de Dados (PGD) previsto para o PMBA fornecerá as orientações para o armazenamento dos dados brutos e sua disponibilização em repositórios de dados e bancos de dados para atendimento às necessidades internas do Programa e posterior transição para a Fundação Renova.

Resumidamente, o PGD estabelece uma metodologia para que os dados sejam mantidos com segurança, e estejam disponíveis para as gerações de tomadores de decisão e pesquisadores ao longo do tempo. Também são detalhados os mecanismos de gestão de dados, com especificações sobre:

- Repositórios para armazenamento dos dados.
- Metadados para descrição dos dados.
- Banco de Dados dirigido às questões/temáticas cobertas pelo PMBA.
- Formas de visualização dos dados.
- Infraestrutura física e humana necessária para a implementação do Plano e manutenção dos dados a longo prazo.

O plano foi baseado no modelo "Digital Curation Centre" fornecido por Digital Curation Centre (DCC), acessado em 01/04/2019. Maiores informações podem ser consultadas na URL a seguir:

<http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>

Abaixo, os principais tópicos abordados no decorrer deste Plano.

- **Arquitetura de Dados:** Apresentação das camadas de dados e os respectivos requisitos de hardware e software.
- **Coleta de Dados:** Apresentação dos tipos de dados coletados e criados e como são coletados e criados.
- **Documentação e Metadados:** Apresentação dos padrões de documentação e de metadados a serem utilizados.
- **Ética e Direitos Autorais:** Apresentação de como serão administradas as questões éticas e geridos os direitos autorais.
- **Armazenamento e Backup:** Apresentação das políticas de armazenamento e backup, além de gestão de acessos e segurança da informação.
- **Seleção e Preservação:** Apresentação do plano de preservação a longo prazo dos conjuntos de dados (dataset).
- **Compartilhamento de Dados:** Apresentação das regras e restrições de compartilhamento de dados.
- **Responsabilidades e Recursos:** Apresentação dos recursos, papéis e responsabilidades relacionados ao gerenciamento dos dados.
- **Integração UGD e UAIR:** Apresentação detalhada dos pontos de integração de ações e produtos entre as Unidades de Gestão de Dados (UGD) e de Atuação Integrada em Rede (UAIR).
- **Custos de Implementação:** Apresentação dos custos de infraestrutura (hardware e software) e de recursos humanos necessários para implementação e manutenção do gerenciamento de dados proposto no Plano de Gestão de Dados.
- **Cronograma de Implementação:** Apresentação de proposta de cronograma para implementação das camadas de dados e processos de gestão definidos no Plano de Gestão de Dados.

2 ARQUITETURA DE DADOS

Nesta seção é apresentado o desenho de arquitetura de dados definido para o Programa de Monitoramento da Biodiversidade Aquática (PMBA). São detalhadas todas as camadas de dados da arquitetura e apresentados os requisitos de hardware e software para as camadas sob gestão da Unidade de Gestão de Dados (UGD).

- Disponibilidade dos dados.
- Frequência de atualização (online, mensal, trimestral).
- Nível de organização e padronização.
- Nível de detalhe dos dados (nível de agregação).
- Tipo de dados (bruto, primário, secundário).
- Finalidade de uso (acompanhamento, divulgação, análises).
- Público alvo (pesquisadores, instituições, sociedade).

Para visualização da apresentação com a solução de dados do PMBA, favor consultar o anexo 22 (Apresentação da Solução de Dados), disponibilizado na seção ANEXOS.

A seguir, serão detalhadas cada uma das camadas, as ferramentas utilizadas e a sua importância para o projeto.

2.1.1 Origem de Dados

Os monitoramentos são feitos manualmente, com sensores ou outros instrumentos e os dados, posteriormente, são colocados em formato digital. Todos os equipamentos e sistemas informacionais utilizados neste processo estão na camada de origem dos dados. Vale informar que os monitoramentos de campo são feitos em terra e/ou embarcado a partir de instrumentação e metodologia específicas de cada Anexo, segundo o Protocolo de Amostragem.

Para o PMBA alguns sistemas informacionais foram contratados para apoiar a digitalização destas informações, entre eles:

- ALLIMS: Sistema de gerenciamento de amostras laboratoriais que centraliza todas as amostras coletadas em campo e é responsável pela rotina analítica dos laboratórios, gerenciando estoques de reagentes e soluções, data de validade de produtos e recursos necessários ao contínuo andamento do Programa de Monitoramento. A ideia é que o ALLIMS possa gerar um banco de dados bem abrangente para que outros sistemas possam buscar e integrar essas informações.
- Primavera: Trata-se da melhor ferramenta do mercado em gestão de projetos. Com o sistema, é possível gerenciar as atividades do Programa de Monitoramento e também seus recursos e materiais. A escolha pelo software se deu por sua capacidade de trabalhar todas as áreas do Programa de forma individual e ao mesmo tempo coletiva, facilitando a entrega de resultados e alimentação de dados.
- Pacote MS Office: Suíte de aplicativos para escritório que contém programas como: processador de texto, planilha de cálculo, banco de dados, apresentação gráfica, cliente de e-mails, entre outros

Todos os sistemas informacionais citados acima e outros referentes a camada de Origem de Dados já foram adquiridos e implantados utilizando infraestrutura própria pela área Escritório de Projetos

(ESCPRO) do PMBA. A Unidade de Gestão de Dados (UGD) não participou do processo de contratação e implementação destas soluções.

2.1.2 Armazém de Dados

Camada de armazenamento e sincronização de arquivos em nuvem. Permite o armazenamento de fotos, documentos, vídeos e outros arquivos, e garante um backup automático dos dados com uma área ilimitada na nuvem. Por exemplo, se o computador pessoal ou smartphone de um membro da equipe for danificado ou perdido, ainda assim, os arquivos permanecem seguros na nuvem (internet).

Para o PMBA, foi contratada a ferramenta Dropbox Business, com os seguintes objetivos:

- Prover espaço de armazenamento ilimitado para os dados gerados ao longo do projeto.
- Permitir o acesso aos arquivos do projeto a qualquer hora e em qualquer lugar, a partir de acesso à internet.
- Possibilitar a recuperação de arquivos via histórico de versões.
- Simplificar o compartilhamento de dados entre equipes, com segurança e evitando o envio de dados por e-mails e outras ferramentas não especializadas (whatsapp, skype, etc.).
- Garantir maior segurança no acesso e compartilhamento de dados.
- Possibilitar a sincronização de dados entre os variados dispositivos da equipe, incluindo computadores, smartphone e tablets.
- Padronização e organização de folders/pastas de equipe para os subprojetos.

O Dropbox deve ser utilizado para armazenar qualquer tipo de dados, em qualquer formato, seja ele temporário ou permanente (oficial). A ideia é que a camada possa ser utilizada como uma área de armazenamento seguro para todos os arquivos e informações digitais gerados pelo PMBA. Por ser uma das primeiras camadas de gestão de dados, os dados disponibilizados estarão em seu formato mais bruto, com poucos ou nenhum tratamento e sem os metadados relacionados (informações sobre os dados). Porém, os dados estarão disponíveis assim que gerados, possibilitando que os usuários já trabalhem sobre eles para a geração de relatórios e outras informações secundárias.

Para a escolha da solução de Armazém de Dados foram avaliadas diversas ferramentas disponíveis no mercado. Dentre elas, o Dropbox Business se mostrou a opção mais equilibrada considerando preço, armazenamento, segurança e sistemas operacionais habilitados. A possibilidade de utilização da solução com sistemas Linux é uma grande vantagem competitiva. Os ambientes Linux são reconhecidos por sua grande segurança e não proliferação de vírus computacionais, além disso são sistemas operacionais frequentemente utilizados para servidores de repositórios de arquivos, bancos de dados e sistemas web e já estão previstos estes equipamentos na infraestrutura do PMBA. A possibilidade de compartilhamento de arquivos com outros usuários que não estejam utilizando o plano

Business torna o seu preço final bem atrativo, pois permite a contratação de licenças para um menor número de usuários se comparado a outros concorrentes.

Para visualização da apresentação da camada de Armazém de Dados do PMBA, favor consultar o anexo 23 (Apresentação do Armazém de Dados), disponibilizado na seção ANEXOS. Na apresentação é possível visualizar maiores detalhes sobre a organização dos grupos de usuários e estrutura de pastas da ferramenta. A Unidade de Gestão de Dados (UGD) também já disponibilizou no anexo 26, na seção ANEXOS, um Guia de Utilização do Dropbox, documentando os passos necessários para configurar e iniciar o uso da solução.

O Dropbox Business foi adquirido e implantado pela Unidade de Gestão de Dados (UGD). Maiores detalhes sobre os requisitos de hardware e software, além de melhores práticas de utilização, serão comentados ao longo do documento.

2.1.3 Repositório de Dados

São áreas de armazenamento online que reúnem de maneira organizada a produção científica de uma instituição ou área temática, armazenando metadados e dados de diferentes formatos.

Para ser útil para cientistas, profissionais envolvidos com o manejo ambiental e formuladores de políticas públicas, um conjunto de dados deve ser devidamente armazenado em uma estrutura organizada e sempre vir acompanhado de explicações detalhadas, incluindo informações sobre como foram coletados, onde, por quem e quando, os chamados Metadados Primários.

Com isso, assegura-se que esse conjunto de dados possa ser entendido e utilizado futuramente por qualquer pessoa, não necessariamente por quem os coletou. Os repositórios de dados resultam em uma série de benefícios tanto para os pesquisadores quanto às instituições ou sociedades científicas, proporcionam maior visibilidade aos resultados de pesquisas e possibilitam a preservação da memória científica de sua instituição.

O programa Morpho deverá ser utilizado para criar, editar e armazenar os metadados. O software Metacat é responsável pela publicação, busca, visualização e download online dos data sets.

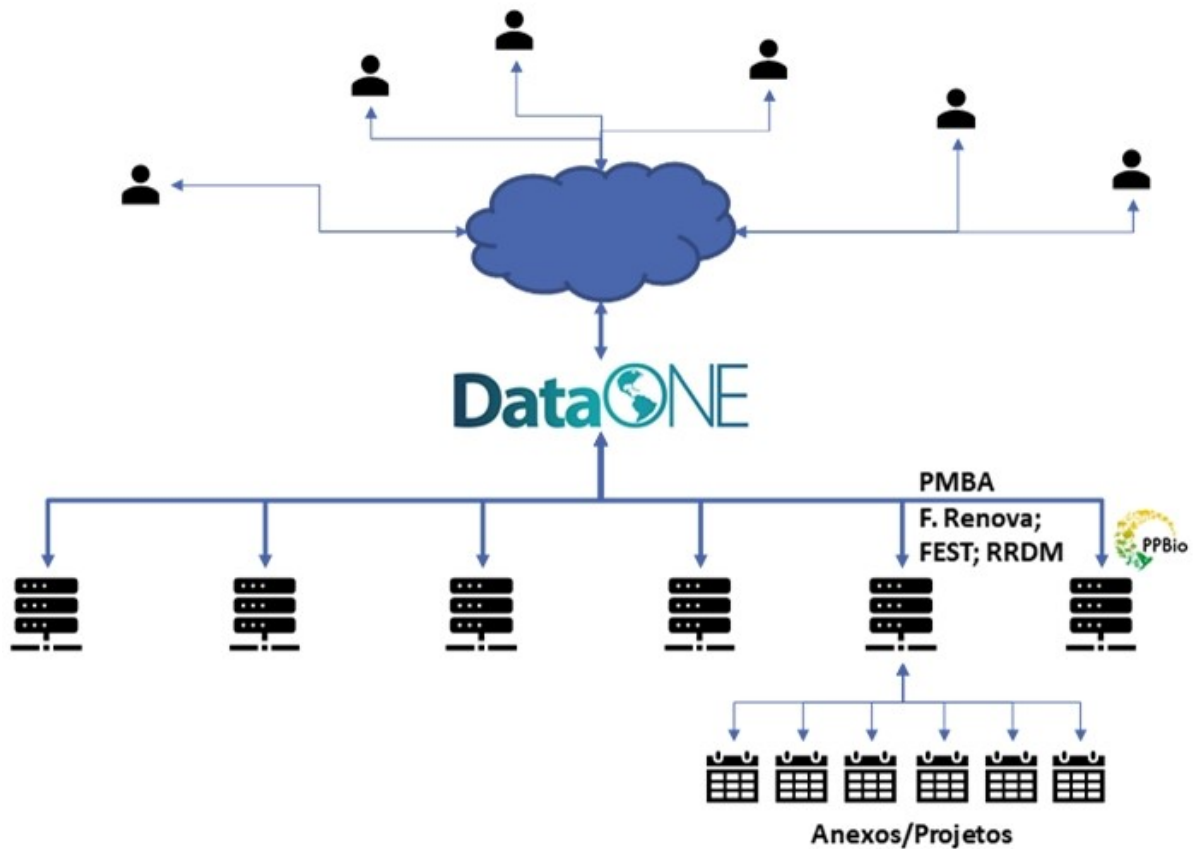
Como parte do Plano de Gestão de Dados (PGD), o repositório de dados do projeto será integrado como um nó membro do DataONE. O DataONE é uma rede internacional de repositórios de dados contendo dados de observações da Terra, incluindo dados de campos como ecologia, biologia, evolução e ciências ambientais, como hidrologia, oceanografia e ciência atmosférica. A rede conta com a participação de centenas de estações de campo, universidades e agências governamentais como nós membros do DataONE.

São inúmeras as vantagens de se tornar um nó membro desta rede de repositórios de dados:

- Alcance a um maior público: As coletas de dados expostas por meio da rede internacional de biodiversidade são detectáveis por um público mais amplo e prontas para uso por um número maior de ferramentas analíticas.
- Aproveitamento da infraestrutura: A infraestrutura atual da rede de biodiversidade suporta ativamente as ferramentas de dados existentes, bem como novas ferramentas a serem construídas para interagir com objetos de dados registrados pelos nós membros. A rede já disponibiliza um portal na Web que pode ser personalizado para cada membro, permitindo consultas e o gerenciamento dos dados, evitando a necessidade de desenvolvimentos adicionais de portais na internet específicos para cada projeto e organização.
- Reconhecimento e crédito: Os produtos de dados publicados são facilmente citáveis e, com uma exposição mais ampla, é possível aumentar o valor do trabalho dos projetos e organizações membros recebendo crédito por meio de citações de dados na literatura publicada.
- Alta disponibilidade: A rede de biodiversidade internacional tem a capacidade de replicar o conteúdo entre os nós membros. Com réplicas disponíveis em outros locais, os custos associados ao atendimento de uma comunidade mundial são minimizados e a disponibilidade do dado é otimizada.
- Maiores oportunidades de colaboração: A divulgação e compartilhamento dos trabalhos com cientistas e instituições internacionais também pode levar a oportunidades de colaboração no futuro. Simplesmente limitando o acesso aos dados do projeto aos parceiros de colaboração, os membros podem compartilhar dados do projeto sem a necessidade de arranjos de compartilhamento de dados caso a caso.

A Figura 2 apresenta o desenho de arquitetura da rede internacional DataONE, incluindo o PMBA como um dos nós do sistema.

Figura 2: Inclusão do PMBA como um nó da rede internacional.



Fonte: autoria própria.

Diferentemente do Dropbox, o Repositório de Dados deve ser utilizado para armazenar a produção científica de uma instituição ou área temática, ou seja, os dados e metadados oficiais do monitoramento. Arquivos temporários e administrativos não devem fazer parte desta camada. O grande benefício desta camada é disponibilizar os dados de monitoramento enriquecidos com os metadados, permitindo que outros usuários possam encontrar estas informações através de pesquisas utilizando informações relativas aos dados, por exemplo espécies monitoradas, área geográfica, parâmetros de monitoramento, pesquisador responsável, período temporal, métodos utilizados, etc. Devido a necessidade de associação dos metadados aos dados, as informações estarão disponíveis nesta camada com um atraso em relação a sua geração se comparado ao Dropbox.

Para a escolha da solução de Repositório de Dados foram avaliadas diversas opções disponíveis no mercado. Dentre elas, o Morpho (utilizado na geração dos conjuntos de dados com metadados) e o Metacat (utilizado para armazenamento e buscas dos conjuntos de dados) se mostraram a solução mais vantajosa para o PMBA. As duas ferramentas devem ser utilizadas em conjunto. O padrão de metadados utilizado pelos programas (EML - Ecological Metadata Language) é um padrão de metadados desenvolvido para as ciências da terra, ambientais e ecológicas, difundido e utilizado em nível internacional. Os softwares foram desenvolvidos pela rede The Knowledge Network for Biocomplexity (KNB - <http://knb.ecoinformatics.org/index.jsp>), são gratuitos e amplamente utilizados na

comunidade científica (<http://knb.ecoinformatics.org/community.jsp>), permitindo a sua integração com o DataONE, citado anteriormente. Esse sistema nos traz muitas vantagens, dentre elas podemos destacar:

- Melhor qualidade na descrição dos dados, com campos mais detalhados para descrever cada item.
- Maior flexibilidade de permissões nos dados e metadados, podendo armazenar os metadados e dados no servidor, mas escolher para quem deixar disponível, concedendo permissões diferentes para metadados e dados.
- Sistema de busca mais eficiente.
- Maior segurança. Futuramente, o nosso servidor estará conectado ao servidor central do DataONE, o que significa que teremos uma cópia de segurança (um backup) de cada metadado em um local físico diferente.
- Maior divulgação e integração com a comunidade internacional. Futuramente, o nosso servidor estará divulgado no site KNB e no sistema DataONE e nós utilizaremos o mesmo sistema de grandes redes ou projetos ecológicos em nível mundial.

Para visualização da apresentação da camada de Repositório de Dados do PMBA, favor consultar o anexo 24 (Apresentação do Repositório de Dados), disponibilizado na seção ANEXOS. Na apresentação é possível visualizar maiores detalhes sobre cada ferramenta e exemplos de telas das aplicações.

O Repositório de Dados do PMBA ainda não está implantado. A configuração das ferramentas Morpho e Metacat estão previstas para o mês posterior à entrega do Plano de Gestão de Dados. A integração do repositório com a rede internacional de biodiversidade será iniciada na segunda metade do PMBA, mas a sua finalização depende de alinhamentos e homologação pelo DataONE, e poderá não acontecer durante o primeiro ano de monitoramento. Maiores detalhes sobre os requisitos de hardware e software, além de melhores práticas de utilização, serão comentados ao longo do documento.

2.1.4 Banco de Dados

São coleções organizadas de dados que se relacionam de forma a criar algum sentido (Informação) e dar mais eficiência durante uma pesquisa ou estudo. São de vital importância para as organizações, e há duas décadas se tornaram a principal peça dos sistemas de informação.

As bases de dados relacionais permitem aos utilizadores (incluindo programadores) escreverem consultas (queries) que não foram antecipadas por quem projetou a base de dados. Como resultado, bases de dados relacionais podem ser utilizadas por várias aplicações em formas que os projetistas originais não previram, o que é especialmente importante em bases de dados que podem ser utilizadas durante décadas ou longos períodos.

Os Sistemas Gerenciadores de Bancos de Dados Relacionais (SGBDR) são o método preferido de armazenamento/recuperação de dados/informações para aplicações multiusuárias grandes onde a coordenação entre muitos usuários é necessária, por exemplo para análises integradas de impacto utilizando ferramentas analíticas, como o Tableau.

Para a escolha da solução de Banco de Dados foram avaliadas diversas ferramentas disponíveis no mercado. Dentre elas, o SGBDR PostgreSQL se mostrou a opção mais atrativa considerando preço, armazenamento, segurança e sistemas operacionais habilitados. O PostgreSQL é um dos SGBDR de código aberto mais avançados existentes no mercado. É de fácil instalação, possui interfaces simples e intuitivas. Por ser leve, não demora mais do que alguns minutos para instalar, mostrando a flexibilidade e eficiência da ferramenta. Possui mecanismos de bloqueio, suporta tamanhos ilimitados de linhas, bancos de dados e tabelas, concede segurança contra falhas (caso tenha um pique de luz, por exemplo), entre tantas outras qualidades. A solução também é multiplataforma e altamente escalável. Além disso, todas as funcionalidades do PostgreSQL estão disponíveis gratuitamente.

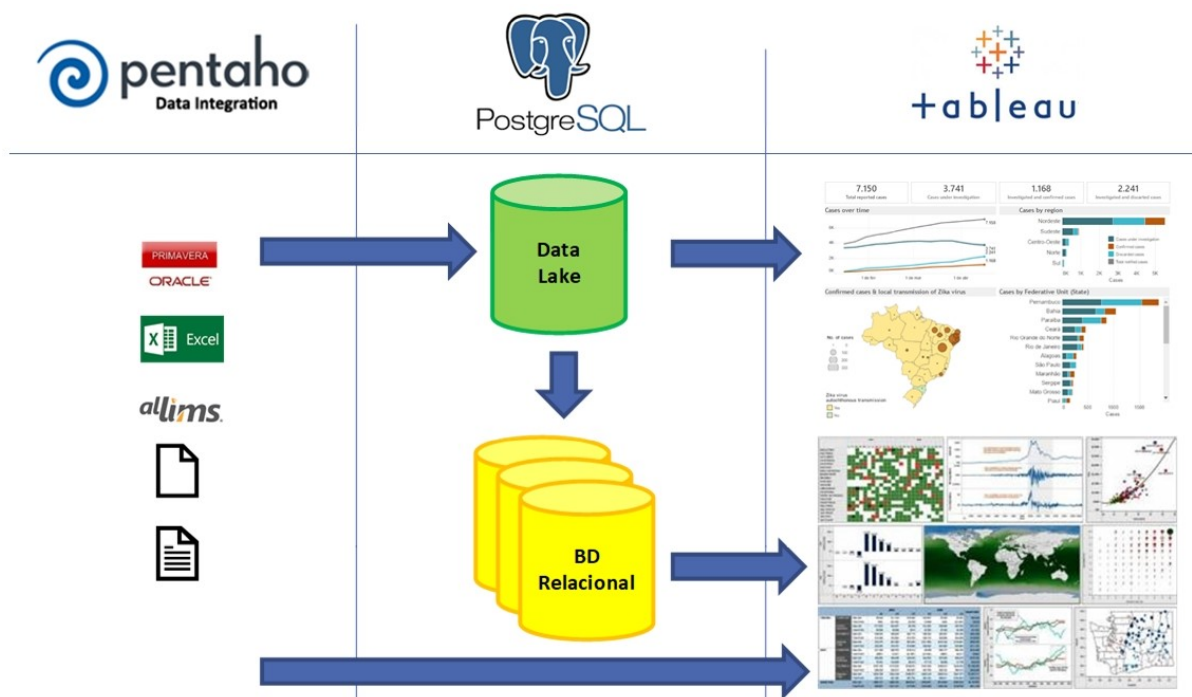
Dentre as camadas de dados apresentadas, o Banco de Dados é a que exige maior tempo para implementação devido à complexidade dos relacionamentos necessários entre as informações. Diferentemente das outras camadas de dados, o banco de dados exigirá uma equipe de desenvolvimento para modelagem e carga das estruturas de dados a serem disponibilizadas. Em função disso e visando entregar um primeiro conjunto de dados no PostgreSQL para apoiar as análises integradas previstas para o primeiro ano de monitoramento, o Banco de Dados do PMBA foi subdividido nas duas subcamadas a seguir:

Data Lake: Camada de dados mais simples, ainda sem todos os relacionamentos necessários entre as informações. O principal objetivo da camada é disponibilizar todos os dados do projeto em um ambiente único de banco de dados para início da exploração. A camada também será utilizada como apoio para definir a modelagem dos dados da camada final de banco de dados.

Base de Dados Relacional: Camada de dados final e mais complexa, com todos os relacionamentos necessários entre as informações. O principal objetivo da camada é disponibilizar dados com qualidade para as ferramentas analíticas, como o Tableau.

A figura abaixo (Figura 3) apresenta o desenho de arquitetura da camada de Banco de Dados.

Figura 3: Arquitetura da camada de Banco de Dados, subcamadas e ferramentas



Fonte: autoria própria.

A subcamada de Data Lake é responsável por receber os dados das origens, Dropbox e Repositório de Dados e já permite que os dados sejam acessados pelas ferramentas analíticas, como o Tableau. O Data Lake também será a origem dos dados para a Base de Dados Relacional. Conforme apresentado na Figura 3, a ferramenta computacional gratuita Pentaho Data Integration (PDI) será utilizada para carga dos dados no PostgreSQL. A ferramenta Tableau é a ferramenta analítica oficial do PMBA para consumo dos dados e geração das análises integradas.

Diferentemente das camadas anteriores, o Banco de Dados deve ser utilizado principalmente para a realização de análises de dados utilizando ferramentas analíticas, como Tableau. A grande vantagem desta camada é o armazenamento dos dados de monitoramento com maior qualidade e relacionamentos pré-definidos entre as informações facilitando a geração de análises integradas entre assuntos. Devido a necessidade de processos de melhoria da qualidade de dados e de processos que garantam o relacionamento das informações, os dados estarão disponíveis nesta camada com um atraso em relação às camadas predecessoras.

Para visualização da apresentação da camada de Banco de Dados do PMBA, favor consultar o anexo 25 (Apresentação do Banco de Dados), disponibilizado na seção ANEXOS.

O Banco de Dados do PMBA ainda não está implantado. A implementação do banco de dados PostgreSQL está programada para iniciar na segunda metade do primeiro ano de monitoramento. A camada de Data Lake, por ser mais simples, e exigir menor esforço de implementação, será entregue ainda durante o primeiro ano do PMBA. Maiores detalhes sobre os requisitos de hardware e software, além de melhores práticas de utilização, serão comentados ao longo do documento.

2.1.5 Análise de Dados

São sistemas de visualização de dados interativos voltados para Business Intelligence (BI). O PMBA contratou o Tableau como sua ferramenta de análise de dados. Trata-se da ferramenta mais utilizada atualmente no mercado, especializada em técnicas de visualização para explorar, integrar e analisar bancos de dados relacionais e cubos de dados.

O Tableau tem uma funcionalidade de mapeamento que é capaz de plotar coordenadas geográficas com latitude e longitude e conectar-se a arquivos espaciais como Esri Shapefiles, KML e GeoJSON para exibir dados geográficos de forma personalizada e integrada com outras informações. É possível agrupar geografias para criar visões e territórios personalizados.

A ferramenta será intensamente utilizada para geração das análises integradas de impacto previstas para o projeto.

O Tableau já foi adquirido e está sendo implantado utilizando infraestrutura própria definida pela área Escritório de Projetos (ESCPRO) do PMBA. A Unidade de Gestão de Dados (UGD) não participou do processo de contratação e implementação desta solução.

2.2 REQUISITOS DE SOFTWARE

Os requisitos de software, ferramentas computacionais necessárias para implementação das camadas de gestão de dados do PMBA são apresentadas a seguir para as camadas sob gestão da Unidade de Gestão de Dados (UGD).

2.2.1 Armazém de Dados

A solução utilizada para a camada de Armazém de Dados é o Dropbox Business. Trata-se de uma ferramenta de armazenamento e sincronização de arquivos em nuvem. Portanto, a infraestrutura de hardware e software é mantida na nuvem pela empresa fornecedora da solução. E este serviço deve ser contratado anualmente pelo PMBA enquanto durar o projeto.

Os usuários conseguem utilizar o serviço e acessar seus dados das três formas abaixo:

- Acesso pela web através de browsers como Google Chrome ou Internet Explorer. Neste caso, não é necessário instalar nenhuma aplicação do Dropbox, basta acessar o site “www.Dropbox.com”, se autenticar e interagir com seus dados.
- Acesso pelo seu computador através do aplicativo Dropbox, onde os dados serão sincronizados com pastas da sua estação de trabalho. É possível cadastrar ilimitados computadores com a mesma conta do Dropbox Business.

- Acesso por smartphones ou tablets através do aplicativo Dropbox. O aplicativo permite a pré-visualização, edição e compartilhamento de arquivos diretamente do smartphone. É possível cadastrar ilimitados equipamentos com a mesma conta do Dropbox Business.

Para a primeira opção (acesso via web), o Dropbox recomenda a utilização das versões mais recentes destes navegadores (browsers) abaixo:

- Chrome
- Firefox
- Internet Explorer 11
- Microsoft Edge
- Safari

Para as segunda e terceira opções, é necessário instalar o aplicativo do Dropbox no dispositivo. Abaixo, são apresentados os requisitos de sistema para os diversos dispositivos e sistemas operacionais disponíveis.

- Computador Mac (Apple):

- Sistema operacional OS X Yosemite (10.10) até macOS Mojave (10.14).
- A versão mais recente do aplicativo do Dropbox para computadores Mac.

- iPhone, iPad ou iPod touch (iOS):

- iOS (sistema operacional) 10 ou superior.
- O aplicativo do Dropbox para iOS mais recente.

- Windows 10 em modo S (Plataforma Universal do Windows para computadores, dispositivos móveis, tablets, Xbox, etc.):

- O sistema operacional Windows 10 com Windows 10 Creator Update (15063) e superior.
- O aplicativo mais recente do Dropbox para Plataforma Universal do Windows.

- Computador Windows (Windows 7, 8, 8.1, 10 não no modo S):

- Sistema operacional Windows 7, 8, 8.1 ou 10 não no modo S.
- A versão mais recente do aplicativo do Dropbox para Windows (não no modo S).

- Dispositivo Android:

- Sistema operacional Android 4.4 (KitKat) ou superior.
- A versão mais recente do aplicativo do Dropbox para Android.

- Computador Linux:

- Sistema operacional Ubuntu 14.04 ou superior, Fedora 21 ou superior.
- Glibc 2.19 ou superior.
- O aplicativo mais recente do Dropbox para Linux.
- Utilização de um dos ambientes desktop a seguir:
 - GNOME shell.
 - Unity.
 - GNOME Classic.
 - XFCE.

A Unidade de Gestão de Dados (UGD) também já disponibilizou no anexo 26, na seção ANEXOS, um Guia de Utilização do Dropbox, documentando os passos necessários para configurar e iniciar o uso da solução. Este guia foi enviado a todos os coordenadores e pesquisadores.

2.2.2 Repositório de Dados

Como comentado anteriormente, para a solução de repositório de dados serão utilizadas as ferramentas Morpho (para criar, editar e armazenar os metadados) e Metacat (responsável pela publicação, busca, visualização e download online dos data sets).

O software Morpho deve ser instalado nos computadores pessoais dos pesquisadores e outros usuários responsáveis pela criação dos data sets (conjuntos de dados englobando metadados e arquivos de dados). O Morpho permite ao usuário criar um catálogo local de dados e metadados que podem ser consultados, editados e visualizados. O Morpho também faz interface com o servidor Metacat, Knowledge Network for Biocomplexity (KNB), que permite aos cientistas fazer upload, download, armazenamento, consulta e visualização de metadados e dados públicos.

O Morpho está disponível para as seguintes plataformas: Linux; Windows; e Mac. Como pré-requisito, é necessário que o sistema possua o Java 1.7 ou mais recente instalado.

O Metacat é uma aplicação servidora, portanto não exige instalação nos computadores pessoais. Porém, exige a instalação em um servidor de dados. Abaixo, são apresentados os requisitos mínimos de sistema para instalação do Metacat em um servidor de dados.

- Sistema operacional Linux.
- Banco de Dados PostgreSQL.
- Apache Jakarta-Ant.
- Apache Jakarta-Tomcat.
- Apache Web Server.
- Java 6.

2.2.3 Banco de Dados

O SGBDR a ser utilizado pelo projeto será o PostgreSQL. Trata-se de uma aplicação servidora e por isso não exige nenhuma instalação nos computadores pessoais dos integrantes do PMBA.

O PostgreSQL é compatível com ACID desde 2001 e possui poderosos complementos, como o popular extensor de banco de dados espaciais PostGISgeospacial. Além disso, pode ser instalado e executado em todos os principais sistemas operacionais: BSD; FreeBSD; OpenBSD; Linux; macOS; Solaris; e Windows.

Como ferramenta auxiliar ao Banco de Dados PostgreSQL, deverá ser instalado, também no ambiente servidor, o Pentaho Data Integration (PDI). O PDI é um software de código aberto desenvolvido em Java, especialista em processos de ETL (Extraction, Transformation and Load) e será utilizado para carga de dados nas tabelas PostgreSQL. O PDI também pode ser instalado nos principais sistemas operacionais: Linux; Windows; e Mac OS.

Como os servidores utilizados serão compartilhados com as outras camadas de dados (armazém de dados e repositório de dados), a recomendação é pela utilização do Sistema Operacional Linux. O sistema operacional é frequentemente utilizado para servidores de repositórios de arquivos, bancos de dados e sistemas web por sua reconhecida segurança e não proliferação de vírus computacionais.

2.3 REQUISITOS DE HARDWARE

Os requisitos de hardware, equipamentos e outros dispositivos computacionais necessários para implementação das camadas de gestão de dados do PMBA são apresentadas a seguir para as camadas sob gestão da Unidade de Gestão de Dados (UGD).

2.3.1 Armazém de Dados

Como já comentado, a infraestrutura de hardware e software do Dropbox Business (solução utilizada para a camada de Armazém de Dados) é mantida na nuvem pela empresa fornecedora da solução e o serviço deve ser contratado anualmente pelo PMBA enquanto durar o projeto para que se tenha acesso aos dados.

Para que os usuários possam acessar os arquivos utilizando o aplicativo Dropbox, são necessários os seguintes requisitos mínimos de hardware:

- 500 MB de espaço no disco rígido.
- 2 GB de RAM.

O espaço em disco apresentado é o mínimo exigido para funcionamento da aplicação. Mas é importante frisar que o usuário que deseja manter uma cópia dos seus dados localmente em seu computador precisará de um espaço em disco equivalente a volumetria de dados que se deseja sincronizar.

O contrato com o Dropbox prevê armazenamento ilimitado na nuvem, mas a Unidade de Gestão de Dados (UGD), por questões de segurança, recomenda que todo o conteúdo seja sincronizado para um servidor de arquivos local e para outro servidor de arquivos na nuvem (servidor independente do fornecedor do Dropbox). Estes servidores serão compartilhados com as outras camadas de dados (repositório de dados e banco de dados) a serem disponibilizadas no PMBA, porém em função da grande quantidade de dados do Programa, foi necessário estimar a volumetria necessária em discos para armazenamento do conteúdo.

Dentre as atividades da UGD, foi realizado um levantamento das volumetrias de dados esperadas para cada um dos Anexos/Projetos do PMBA através de reuniões com os seus coordenadores. Abaixo, é apresentado o levantamento, considerando os 5 anos de projeto (Quadro 2).

Quadro 2: Levantamento de volumetria de dados para cada um dos Anexos/Projetos.

Anexo	Descrição Volumetria
1- ECOTOXICOLOGIA	As planilhas Excel costumam ter entre 500Kb e 1Mb.
1- ECOTOXICOLOGIA	Os gráficos possuem entre 40Kb e 50Kb.
2 3-DULCÍCOLA-INTEGRAÇÃO	Foi comentado que o equivalente os 2 meses de dados brutos gerou um conjunto de arquivos ainda na ordem de grande de Kbytes.
2 3- DULCÍCOLA-INTEGRAÇÃO	Mas também serão utilizados fotos e vídeos, o que demandaria mais espaço.
3- MARINHO-ESTUARINO	A previsão de volumetria é da ordem de dezenas de Terabytes (140TBytes).
3- MARINHO-ESTUARINO	Os dados brutos exigem utilização de softwares pagos para utilização. Dados de modelo que deveriam ser abertos por softwares nas planilhas de dados.
4- PRAIAS	Os arquivos Excel são de pequeno e médio porte apenas com dados não passando de poucos MBs.
5- MANGUEZAL	Também serão gerados muitos arquivos de imagens. Os pesquisadores fotografam muito material em campo e há possibilidade de criação de vídeos também.
5- MANGUEZAL	Também será realizada inserção de banco de dados externos.
5- MANGUEZAL	Os arquivos Excel são de pequeno e médio porte apenas com dados não passando de poucos MBs.
5- RESTINGA	Também são gerados arquivos para acervo fotográfico utilizado principalmente para a taxonomia.
5- RESTINGA	Alguns vídeos de situações encontradas durante as atividades de coleta também serão armazenados.
5- RESTINGA	O Anexo irá gerar um grande volume de imagens e vídeos.
6- MEGAFUNA	Porém, apenas devem ser considerados como dados, a informação retirada dos vídeos e passadas para planilhas.
6- MEGAFUNA	As planilhas geradas possuem dezenas ou no máximo centenas de registros com variáveis que descrevem as ocorrências.
6- MEGAFUNA	Também poderão ser geradas imagens e vídeos com os recortes, criando um grande banco de imagens, inclusive com a intenção de divulgar, posterior.
6- MEGAFUNA	Foi comentado que os documentos gerados são pequenos. Arquivos pequenos ou médios.
7- ICTIOFAUNA DULCÍCOLA	Todos os documentos e outros materiais podem ser manipulados por estações de trabalhos manuais, sendo em geral de pequeno porte.
7- ICTIOFAUNA DULCÍCOLA	As planilhas geradas não devem ultrapassar os 20MBytes.
7- ICTIOFAUNA MARINHA	Serão geradas várias horas de vídeos por campanha. Estão atualmente utilizando uma HD de 2TB, mas a estimativa final para o projeto é de 10TB.

Anexo	Descrição Volumetria
7- ICTIOFAUNA MARINHA	Foi comentado que os documentos gerados são relativamente pequenos.
8- SEDIMENTAÇÃO ABROLHOS	As planilhas e imagens geradas podem ser facilmente manipuladas por computadores pessoais e compartilhadas por e-mail, por exemplo.

Fonte: autoria própria.

Com exceção do Anexo/Projeto 3 – Marinho-Estuarino, todos os demais possuem volumetrias na ordem de grandeza de MBytes ou GBytes. O Anexo/Projeto 3 - Marinho-Estuarino levantou a necessidade de dezenas de TBytes para armazenamento de conteúdo.

Considerando uma volumetria máxima de 1 TBytes por Anexo/Projeto somada a necessidade especial do Anexo/Projeto 3 – Marinho-Estuarino, estimamos um volume de até 160 TBytes de dados para armazenamento em nosso servidor de arquivos local e servidor de arquivos na nuvem (servidor independente do fornecedor do Dropbox).

Os 160 TBytes de dados deverão estar armazenados em Network Attached Storage (NAS) com redundância de disco nível RAID 5 (conjunto redundante de discos independentes que visa obter vantagens na utilização de subsistemas de dois ou mais discos) para garantir maior segurança dos dados.

Os custos de aquisição e suporte desta infraestrutura de discos para armazenamento local e na nuvem está detalhado na seção CUSTOS DE IMPLEMENTAÇÃO.

Atualmente, o PMBA conta apenas com área ilimitada em nuvem contratada pelo Dropbox. A implementação das áreas de discos adicionais para sincronização será realizada após a aprovação do Plano de Gestão de Dados (PGD) do PMBA.

2.3.2 Repositório de Dados

Os computadores pessoais dos pesquisadores e outros usuários responsáveis pela criação dos data sets (conjuntos de dados englobando metadados e arquivos de dados) deverão ter a seguinte configuração mínima para instalação da aplicação Morpho:

- 256 MB de memória RAM disponível no sistema. Mais memória RAM é especialmente útil se houver um grande número de pacotes de dados locais, pois os dados locais são armazenados em cache durante inicialização.

- CPU de no mínimo 700MHz.

Para instalação do servidor do Metacat é necessário garantir os requisitos mínimos de hardware exigidos para instalação do sistema operacional Linux e demais aplicações dependentes informadas nos requisitos de software. A seguir, as exigências mínimas de hardware:

- 128 MB de memória RAM disponível no sistema.

- CPU Pentium III ou superior.

- 150 MB de espaço em disco.

A Unidade de Gestão de Dados (UGD), por questões de segurança, recomenda que sejam implementados dois servidores Metacat, sendo um deles local e outro na nuvem, com sincronização de dados entre eles. Estes servidores serão compartilhados com as outras camadas de dados (armazém de dados e banco de dados) a serem disponibilizadas no PMBA.

A quantidade média de acessos simultâneos estimada ao repositório de dados é de 50 acessos (considerando 1 acesso por subprojeto do PMBA). Os volumes trafegados dependem do tamanho dos arquivos de dados disponibilizados juntamente com os metadados, que segundo levantamentos apresentados no Quadro 2, em sua grande maioria terão volumetrias da ordem de grandeza de KBytes ou poucos MBytes. Sendo assim, o servidor deverá estar exposto na internet, com acesso habilitado apenas para os integrantes do PMBA e conexão de no mínimo 10 Mbps para downloads e uploads.

Com relação ao espaço em disco, estimamos uma volumetria máxima de 10 TBytes para o Repositório de Dados, considerando até 1 TBytes por Anexo/Projeto. Este é o volume máximo previsto para os 5 anos de projeto. Importante frisar que apenas devem ser publicados os arquivos de dados e metadados, após curadoria, no repositório de dados. Arquivos de imagens, vídeos, áudios e outros formatos proprietários de dados, salvo algumas raras exceções, deverão ser mantidos apenas no Dropbox, o que reduz em muito a volumetria final do Repositório de Dados.

2.3.3 Banco de Dados

A seguir, são apresentados os requisitos mínimos de hardware para instalação do PostgreSQL:

- CPU de 1 GHz.
- 1 GB de memória RAM.
- 512 MB de espaço em disco (espaço adicional em disco é necessário para os dados).

Para instalação do Pentaho Data Integration (PDI) é necessário garantir os requisitos mínimos de hardware abaixo:

- 4 GB de memória RAM disponível no sistema.
- CPU de 2 GHz.
- 2 GB de espaço em disco.

Seguindo a mesma estratégia definida para as demais camadas de dados, a Unidade de Gestão de Dados (UGD), por questões de segurança, recomenda que sejam implementados dois servidores de Banco de Dados com a instalação do PostgreSQL e PDI, sendo um deles local e outro na nuvem, com sincronização de dados entre eles. Estes servidores serão compartilhados com as outras camadas de dados (armazém de dados e repositório de dados) a serem disponibilizadas no PMBA.

A quantidade média de acessos simultâneos estimada ao banco de dados é a mesma do repositório de dados, 50 acessos (considerando 1 acesso por subprojeto do PMBA). Os volumes trafegados devem ser bem inferiores ao do Repositório de Dados, já que todas as informações são processadas no próprio servidor e apenas os resultados de requisições transferidas pela rede de dados, não havendo transferência de arquivos. Sendo assim, os requisitos mínimos de rede utilizados para o Repositório de Dados também poderiam ser mantidos para o Banco de Dados considerando que o servidor deverá estar exposto na internet, com acesso habilitado apenas para os integrantes do PMBA e conexão de no mínimo 10 Mbps.

Com relação ao espaço em disco, estimamos uma volumetria máxima de 10 TBytes para o Banco de Dados, considerando a área para dados e índices de dados. A estimativa é que seja disponibilizado no Banco de Dados o mesmo conteúdo já disponibilizado no Repositório de Dados, com a inclusão dos relacionamentos entre as informações e implementação dos processos de qualidade de dados. Portanto, foi mantida a estimativa de até 1 TBytes por Anexo/Projeto. Este é o volume máximo previsto para os 5 anos de projeto. Também é importante frisar que apenas devem ser publicados no Banco de Dados os dados com os resultados analíticos do monitoramento. Imagens, vídeos, áudios e outros conteúdos proprietários, salvo algumas raras exceções, deverão ser mantidos apenas no Dropbox, o que reduz em muito a volumetria final da camada.

3 GESTÃO DE DADOS

Nesta seção serão fornecidas as orientações para o armazenamento dos dados e sua disponibilização em repositórios e bancos de dados para atendimento às necessidades internas do Programa e posterior transição para a Fundação Renova.

Serão apresentados diversos itens metodológicos e de melhores práticas para que os dados sejam mantidos com segurança, e estejam disponíveis para as gerações de tomadores de decisão e pesquisadores ao longo do tempo.

3.1 COLETA DE DADOS

Dados de pesquisa consistentes e bem ordenados serão mais fáceis de encontrar, entender e reutilizar. Nesta seção serão abordados os padrões e métodos relevantes para coleta e organização dos dados.

3.1.1 Quais dados serão coletados ou criados?

O PMBA irá gerar e coletar uma grande variedade de dados. Para formular um plano de gerenciamento de dados, é importante categorizar os dados de quatro maneiras: por origem, formato, estabilidade e volume.

Origem dos Dados:

Embora os dados provenham de muitas fontes diferentes, eles podem ser agrupados em quatro categorias principais.

- Observacional:

- Coletado diretamente em campo.
- Geralmente insubstituível e, portanto, o mais importante para proteger.
- Exemplos: Sensor leituras, telemetria, resultados de pesquisas, imagens.

- Experimental:

- Geralmente gerado em laboratório ou sob condições controladas.
- Frequentemente reproduzível, mas pode ser caro ou demorado.
- Exemplos: sequências genéticas, cromatogramas.

- Simulação:

- Gerados a partir de modelagem numérica.
- Prováveis de serem reproduzíveis se o modelo e as entradas forem preservados.
- Exemplos: modelos climáticos e econômicos.

- Derivado / Compilado:

- Gerado a partir de conjuntos de dados existentes.
- Reproduzível, mas pode ser muito caro e demorado.
- Exemplos: texto e dados mineração, banco de dados compilado, modelos 3D.

O Quadro 3 apresenta o resultado do levantamento realizado pela UGD para identificação de todas as ferramentas utilizadas pelos Anexos/Projetos do PMBA para geração e manipulação de dados.

Como visto, no PMBA teremos todas as 4 categorias de dados, gerados em momentos distintos e com regras de disponibilização diferentes em cada uma das camadas da arquitetura de dados.

- Armazém de Dados: Camada será responsável por armazenar os dados gerados em sua forma mais bruta, ou seja, com poucos metadados relacionados e menor qualidade de dados.

- Repositório de Dados: Camada será responsável por publicar os dados gerados, incluindo os metadados associados e seguindo padrões de formatação e regras de qualidade especificados pelo projeto.

- Banco de Dados: Camada será responsável por disponibilizar os dados gerados para acesso pelas ferramentas analíticas e de integração, responsáveis pelo suporte às análises integradas de impacto. A camada garantirá um maior nível de qualidade dos dados e de relacionamento entre os assuntos.

Quadro 3: Levantamento das ferramentas para geração de dados pelos Anexos/Projetos.

Anexo	Ferramenta
1- ECOTOXICOLOGIA	MS Excel
1- ECOTOXICOLOGIA	R
1- ECOTOXICOLOGIA	Statistica
1- ECOTOXICOLOGIA	SW Gráficos
2 3- DULCÍCOLA-INTEGRAÇÃO	MS Excel
2 3- DULCÍCOLA-INTEGRAÇÃO	R
2 3- DULCÍCOLA-INTEGRAÇÃO	Statica
2 3- DULCÍCOLA-INTEGRAÇÃO	SPSS
2 3- DULCÍCOLA-INTEGRAÇÃO	ArcGIS
2 3- DULCÍCOLA-INTEGRAÇÃO	MVSP
2 3- DULCÍCOLA-INTEGRAÇÃO	SigmaPlot
3- MARINHO-ESTUARINO	MS Excel
3- MARINHO-ESTUARINO	Statistica
3- MARINHO-ESTUARINO	R
3- MARINHO-ESTUARINO	MatLab
3- MARINHO-ESTUARINO	ArcGIS
3- MARINHO-ESTUARINO	Primer
3- MARINHO-ESTUARINO	Jump
3- MARINHO-ESTUARINO	Orange
4- PRAIAS	MS Excel
4- PRAIAS	MS Word
5- MANGUEZAL	MS Excel
5- MANGUEZAL	Statistic
5- MANGUEZAL	Statsof
5- MANGUEZAL	R
5- MANGUEZAL	SIG
5- MANGUEZAL	RTK
5- MANGUEZAL	Juno (Trimble)
5- MANGUEZAL	ArcGIS
5- MANGUEZAL	Idrisi
5- MANGUEZAL	Câmera de infra-vermelho acoplada em drone
5- MANGUEZAL	IRGA
5- MANGUEZAL	Fluorômetro
5- RESTINGA	MS Excel
5- RESTINGA	R
6- MEGAFUNA	R
6- MEGAFUNA	Octave
6- MEGAFUNA	Raven
7- ICTIOFAUNA DULCÍCOLA	MS Excel
7- ICTIOFAUNA DULCÍCOLA	PAST

Anexo	Ferramenta
7- ICTIOFAUNA DULCÍCOLA	Geneious
7- ICTIOFAUNA DULCÍCOLA	Popgene
7- ICTIOFAUNA DULCÍCOLA	Mega
7- ICTIOFAUNA DULCÍCOLA	Arlequin
7- ICTIOFAUNA DULCÍCOLA	Bottleneck
7- ICTIOFAUNA MARINHA	MS Excel
7- ICTIOFAUNA MARINHA	R
7- ICTIOFAUNA MARINHA	Statistica
7- ICTIOFAUNA MARINHA	Prime
7- ICTIOFAUNA MARINHA	ArcGIS
7- ICTIOFAUNA MARINHA	Photoshop
7- ICTIOFAUNA MARINHA	Illustrator
8- SEDIMENTAÇÃO ABROLHOS	ArcMap
8- SEDIMENTAÇÃO ABROLHOS	ArcView

Fonte: autoria própria.

Formato dos Dados:

Em função da abrangência e da grande quantidade de assuntos tratados e monitorados pelo PMBA, a diversidade de formatos de dados gerados também é muito grande. O quadro a seguir apresenta o resultado do levantamento realizado pela UGD para identificação de todos os formatos de dados que serão gerados e manipulados pelos Anexos/Projetos do PMBA (Quadro 4).

Quadro 4: Levantamento dos formatos de dados a serem gerados pelos Anexos/Projetos

Anexo	Formatos de Dados
1- ECOTOXICOLOGIA	Planilhas MS Excel
1- ECOTOXICOLOGIA	Gráficos diversos
1- ECOTOXICOLOGIA	Análises Estatísticas
2 3- DULCÍCOLA-INTEGRAÇÃO	Planilhas MS Excel
2 3- DULCÍCOLA-INTEGRAÇÃO	Fotos
2 3- DULCÍCOLA-INTEGRAÇÃO	Vídeos
2 3- DULCÍCOLA-INTEGRAÇÃO	Mapas
3- MARINHO-ESTUARINO	Planilhas numéricas MS Excel
3- MARINHO-ESTUARINO	Fotos
3- MARINHO-ESTUARINO	Vídeos
3- MARINHO-ESTUARINO	Imagens de Satélite
4- PRAIAS	Planilhas MS Excel
4- PRAIAS	Documentos MS Word
4- PRAIAS	Gráficos diversos
4- PRAIAS	Mapas
5- MANGUEZAL	Planilhas MS Excel
5- MANGUEZAL	Gráficos diversos

Anexo	Formatos de Dados
5- MANGUEZAL	Análises Estatísticas
5- RESTINGAS	Planilhas MS Excel
5- RESTINGAS	Gráficos diversos
5- RESTINGAS	Análises Estatísticas
6- MEGAFUNA	Planilhas
6- MEGAFUNA	Fotos
6- MEGAFUNA	Vídeos
6- MEGAFUNA	Sons
7- ICTIOFAUNA DULCÍCOLA	Parâmetros ecológicos
7- ICTIOFAUNA DULCÍCOLA	Planilhas MS Excel
7- ICTIOFAUNA DULCÍCOLA	Alinhamento de sequências genéticas
7- ICTIOFAUNA MARINHA	Planilhas Excel
7- ICTIOFAUNA MARINHA	Análises Estatísticas
7- ICTIOFAUNA MARINHA	Mapas
7- ICTIOFAUNA MARINHA	Fotos
7- ICTIOFAUNA MARINHA	Vídeos
8- SEDIMENTAÇÃO ABROLHOS	Planilhas MS Excel
8- SEDIMENTAÇÃO ABROLHOS	Imagens
8- SEDIMENTAÇÃO ABROLHOS	Arquivos Geográficos

Fonte: autoria própria.

Como visto acima, os dados do PMBA podem ser apresentados de várias formas, incluindo:

- Textos: notas de campo ou de laboratório, respostas a questionários.
- Planilhas numéricas: tabelas, contagens, medidas.
- Audiovisual: imagens, gravações de som, vídeos.
- Modelos, código de computador.
- Disciplina específica: saídas de equipamentos especialistas.

O formato de arquivo escolhido para os dados é um fator primordial na capacidade de outra pessoa de acessá-lo no futuro. É necessário pensar cuidadosamente sobre qual formato de arquivo será melhor para gerenciar, compartilhar e preservar os dados. As tecnologias mudam continuamente e todos os hardwares e softwares contemporâneos devem se tornar obsoletos em pouco tempo. Logo, é muito importante considerar como os dados serão lidos se o software usado para produzi-lo ficar indisponível. Embora qualquer formato de arquivo escolhido hoje possa se tornar ilegível no futuro, alguns formatos têm maior probabilidade de serem lidos do que outros. Formatos que provavelmente estarão acessíveis no futuro são:

- Formatos de arquivos não proprietários.
- Formatos com padrões documentados.

- Formatos de uso comum pela comunidade de pesquisa.

Formatos que utilizam codificações de caracteres padrão (ou seja, ASCII, UTF-8)

- Formatos descomprimidos/descompactados.

Abaixo, as opções de formato de dados que deverão ser utilizados pelo PMBA sempre que possível. Estes são os formatos de dados oficiais para o PMBA e devem ser respeitados para as publicações no repositório de Dados.

- Imagem: JPEG, JPG-2000, PNG, TIFF.

- Texto: texto simples (TXT), HTML, XML, PDF/A.

- Áudio: AIFF, WAVE.

- Contêineres de compactação: TAR, GZIP, ZIP.

- Bancos de dados: prefira XML ou CSV para formatos binários nativos.

Se o pesquisador julgar necessário ou conveniente trabalhar com dados em um formato de arquivo proprietário/desencorajado, é possível fazê-lo, mas deve considerar salvar o trabalho também em um dos formatos acima assim que finalizadas as atividades sobre os dados e antes de publicá-lo no repositório de dados.

Outros casos, que não possam ser convertidos para os formatos oficiais apresentados anteriormente, deverão ser justificados durante a geração dos metadados para publicação no Repositório de Dados. Também devem ser informados os softwares e versões de softwares necessários, além de outros pré-requisitos para visualização e manipulação dos conteúdos.

Os dados tabulares merecem uma menção especial porque são muito comuns entre disciplinas, principalmente como planilhas do Excel. Além disso, no PMBA foram especificadas planilhas pré-formatadas para divulgação dos resultados de cada parâmetro monitorado. Estas planilhas têm sido utilizadas na geração dos relatórios trimestrais entregues à Fundação Renova e também como divulgação e compartilhamento de dados entre os pesquisadores enquanto a solução de Repositório de Dados não está disponível. Porém, para que os dados destas e de outras planilhas sejam publicados no Repositório de Dados, devem ser tomadas as precauções abaixo.

- Usar o comando "Salvar como ..." para exportar seu trabalho para o formato .csv quando terminar as análises.

- Os dados das planilhas serão mais fáceis de entender e de serem exportados se você seguir as melhores práticas ao configurá-las, como:

- Não colocar mais de uma tabela em uma planilha. Ou, caso exista mais de uma tabela por planilha, separar em abas e exportar cada uma das abas separadamente.
- Incluir uma linha de cabeçalho com título compreensível para cada coluna.

- Separar os gráficos em novas planilhas ou novas abas da planilha, não as incorpore junto com as tabelas de dados.

No anexo 30 (Planilha de Dados Padrão), na seção ANEXOS, é apresentado um exemplo de planilha pré-formatada para divulgação dos resultados acompanhada de instruções para preenchimento.

Estabilidade dos Dados:

Os dados também podem ser corrigidos ou alterados ao longo do projeto (e mesmo além do final do projeto). Com relação a estabilidade, as categorias mais comuns do conjunto de dados são:

- Conjuntos de dados fixos: nunca são alterados após serem coletados ou gerados.
- Conjuntos de dados dinâmicos: novos dados podem ser adicionados e dados antigos podem ser alterados ou excluídos.

A identificação da categoria correta é importante pois afeta como os dados serão organizados e documentados, guiando o nível de versionamento que você precisará realizar.

Para o PMBA, todos os documentos gerados passíveis de alterações deverão ser versionados, incluindo uma nova página (documentos texto como Word), nova aba (documentos tabulares como Excel) ou novo slide (documentos de apresentações como PowerPoint) com o histórico de versionamento, informando a data de alteração, o autor da alteração e uma breve descrição dos itens ajustados ou acrescentados.

No Repositório de Dados, as versões ou alterações de conteúdo são automaticamente geradas a cada nova publicação do data set ou conjunto de dados. Assim, o histórico de versões é armazenado na ferramenta com a data de atualização e o usuário responsável. Mas é importante a inclusão de comentários explicando o conteúdo alterado para que seja possível a visualização pelos consumidores da informação.

Como o projeto possui várias campanhas de monitoramento, os resultados do monitoramento de cada uma das campanhas serão sempre acrescentados a conjuntos de dados já publicados, e a informação dos versionamentos será importante para que os pesquisadores e demais consumidores da informação tenham a abrangência temporal do dado sempre atualizada.

Volumetria dos Dados:

Conforme apresentado anteriormente na seção ARQUITETURA DE DADOS, Quadro 2, foi realizado um levantamento das volumetrias de dados esperadas para cada um dos Anexos/Projetos do PMBA, considerando os 5 anos de projeto. De acordo com o levantamento, a estimativa final para os 5 anos de monitoramento é de 160 TBytes de informações. Esta é a volumetria esperada para o Armazém de Dados (Dropbox), camada que armazenará todo o conteúdo gerado pelo PMBA antes de ser disponibilizado e publicado nas demais camadas de dados (Repositório de Dados e Banco de Dados).

Arquivos de imagens, vídeos, áudios e outros formatos proprietários de dados, salvo algumas raras exceções, deverão ser mantidos apenas no Dropbox, o que reduz em muito a volumetria final das demais camadas de dados (Repositório de Dados e Banco de Dados), conforme já apresentado anteriormente, na seção ARQUITETURA DE DADOS. Como veremos mais adiante, já existe uma estrutura de pastas criadas no Armazém de Dados para organização de arquivos mídias (fotos e vídeos). E devido a flexibilidade da solução Dropbox, é possível sempre referenciar este tipo de conteúdo através de URLs (links da internet), evitando a necessidade de replicar estes arquivos com grandes volumes de dados para as camadas de dados subsequentes.

Novamente, é importante ressaltar que a volumetria estimada só será alcançada no final dos 5 anos de monitoramento, seguindo uma taxa de crescimento contínua e regular ao longo de todo o PMBA. Não são esperados picos de geração de dados ao longo do Programa.

3.1.2 Como os dados serão coletados ou criados?

A seguir serão apresentadas as diversas metodologias de aquisição de dados utilizadas pelo PMBA, bem como regras para organização dos arquivos gerados no Armazém de Dados e procedimentos para garantia mínima da qualidade de dados das publicações em Repositório de Dados.

Metodologias de Coleta de Dados:

Para fins de qualificação e padronização das amostragens em campo, durante os primeiros meses do PMBA, o CTEC em conjunto com os coordenadores de Anexos/Projetos, detalhou os procedimentos de amostragem em campo, acondicionamento e transporte das amostras de água, sedimento e materiais biológicos, bem como de registro de dados in situ a serem seguidos.

Existe uma série de procedimentos padronizados para todos os Anexos/Projetos, porém em função da grande diversidade de assuntos tratados, também foram detalhados procedimentos específicos para cada frente.

Abaixo, alguns procedimentos gerais padronizados para o PMBA e aplicáveis a todos os Anexos/Projetos:

- Procedimentos para Segurança Embarcada e Laboratorial: Atribuições técnicas de embarque e procedimentos de segurança; Relatório de Embarque; Comunicação.
- Procedimentos para Garantia da Qualidade da Amostragem: Coleta, armazenamento, transporte e recebimento de amostras.

Já os Procedimentos para Amostragem em Campo, são específicos de cada Anexo/Projeto.

Todos os procedimentos metodológicos gerais e específicos, citados acima, estão disponíveis para consulta no anexo 27 (Protocolos de Amostragem do CTEC), disponibilizado na seção ANEXOS. Nos anexos 12 a 19 (Planos de Trabalho dos Anexos do PMBA), é possível consultar a metodologia a ser

utilizada para cada um dos Anexos/Projetos, frequências de amostragem, periodicidades das campanhas de monitoramento e os produtos (Dados Brutos e Análise de Dados) a serem gerados.

Ainda na seção ANEXOS, no anexo 28 (Calendário das Campanhas de Coleta), é apresentado o cronograma de realização das 12 campanhas de monitoramento previstas para o primeiro ano do PMBA, com abertura por Anexo/Projeto. E, finalmente, para visualização dos parâmetros a serem monitorados por cada Anexo/Projeto, favor consultar o anexo 29 (Parâmetros do Programa de Monitoramento).

Organização dos Dados:

A primeira camada de dados do PMBA, compartilhada entre todos os pesquisadores, é o Armazém de Dados que utiliza a solução Dropbox. A camada irá receber e armazenar dados de diferentes Anexos/Projetos, campanhas de monitoramento, origens e formatos. Por isso, é muito importante que os dados disponibilizados nesta primeira camada estejam organizados para permitir que todos os usuários consumidores com acesso ao Armazém de Dados possam navegar pelas pastas e informações com facilidade, buscando e encontrando o material desejado com rapidez e evitando acessos frequentes aos gerados do conteúdo para a localização dos arquivos.

O Dropbox licenciado para o projeto permite a ativação de até 13 usuários. Todos os usuários do Dropbox devem utilizar endereços de e-mail do domínio "rrdm.net.br". Abaixo, é apresentada a lista com os usuários ativados (Quadro 5). O usuário do NAIR "nair@rrdm.net.br" também é o Administrador do ambiente.

Quadro 5: Usuários ativados no Dropbox e contatos dos respectivos responsáveis

Usuário DropBox	Anexo/Grupo de Trabalho	Nome Responsável	E-mail Responsável
adaltobianchini@rrdm.net.br	ANEXO 01- ECOTOXICOLOGIA	Adalto Bianchini	adaltobianchini@furg.br
fabian@rrdm.net.br	ANEXO 03-MARINHO-ESTUARINO	Fabian Sá	fabiannetuno@gmail.com
gfbarroso@rrdm.net.br	ANEXO 02_03- DULCICOLA INTEGRAÇÃO	Gilberto Barroso	gfbarroso@gmail.com
albino.jacqueline@rrdm.net.br	ANEXO 04- PRAIAS	Jacqueline Albino	albino.jacqueline@gmail.com
diu@rrdm.net.br	ANEXO 05- RESTINGA	Diolina Moura Silva	diu@terra.com.br
monica.tognella@rrdm.net.br	ANEXO 05- MANGUEZAL	Mônica Tognella	monica.tognella@gmail.com
agnaldo.ufes@rrdm.net.br	ANEXO 06- MEGAFUNA	Agnaldo Silva Martins	agnaldo.ufes@gmail.com
mhostim@rrdm.net.br	ANEXO 07- ICTIOFAUNA MARINHA	Maurício Hostim	mhostim@gmail.com
jdergam@rrdm.net.br	ANEXO 07- ICTIOFAUNA DULCICOLA	Jorge Dergam	jdergam@gmail.com
evangelista.uerj@rrdm.net.br	ANEXO 08- SEDIMENTAÇÃO ABROLHOS	Heitor Evangelista	evangelista.uerj@gmail.com
escritorio@rrdm.net.br	ESCPRO	Eustáquio Vinícius	eustaquiovinicius@uol.com.br

Usuário DropBox	Anexo/Grupo de Trabalho	Nome Responsável	E-mail Responsável
nair@rrdm.net.br	NAIR	Edmilson Teixeira	edmilsonteixeira@hotmail.com
ctec@rrdm.net.br	CTEC	Alex Cardoso Bastos	alexcardosobastos@gmail.com

Fonte: autoria própria.

Também foram criados 14 Grupos de Usuários visando facilitar a concessão de permissões de acesso aos arquivos e pastas criadas no Dropbox. Abaixo, é apresentada a lista com os grupos de usuários criados e os respectivos usuários associados (Quadro 6).

Quadro 6: Grupos de usuários criados no Dropbox e usuários associados

Nome do Grupo	Usuários Associados
00-CTEC	ctec@rrdm.net.br
00-ESCPROJ	eustaquio@rrdm.net.br
00-NAIR	nair@rrdm.net.br
01-ECOTOXICOLOGIA	adaltobianchini@rrdm.net.br
02_03-DULCICOLA_INTEGRACAO	gfbarroso@rrdm.net.br.
03-MARINHO_ESTUARINO	fabian@rrdm.net.br
04-PRAIAS	albino.jacqueline@rrdm.net.br
05-MANGUEZAL	monica.tognella@rrdm.net.br
05-RESTINGA	diu@rrdm.net.br
06-MEGAFUNA	agnaldo.ufes@rrdm.net.br
07-ICTIOFAUNA_DULCICOLA	jdergam@rrdm.net.br
07-ICTIOFAUNA_MARINHA	mhostim@rrdm.net.br
08-SEDIMENTACAO_ABROLHOS	evangelista.uerj@rrdm.net.br
Todos em Rede Rio Doce Mar	Todos os usuários da rede

Fonte: autoria própria

Além do grupo de usuário “Todos em Rede Rio Doce Mar”, também foi criado um grupo de usuário para cada Anexo/Grupo de Trabalho do projeto. Inicialmente, os grupos de usuários foram criados apenas com o usuário responsável pelo Anexo/Grupo de Trabalho do projeto. Posteriormente, outros usuários poderão ser adicionados ao grupo.

Visando uma melhor organização dos dados de cada Anexo/Grupo de Trabalho do PMBA, foram criadas 13 pastas de equipe no Dropbox. Todos os arquivos criados deverão estar localizados abaixo da estrutura de uma das pastas de equipe a seguir (Quadro 7).

Quadro 7: Pastas de equipe e permissões definidas no Armazém de Dados (Dropbox).

Pasta de Equipe	Grupos com Acesso de Visualização	Grupos com Acesso de Escrita
00-CTEC	00-ESCPROJ	00-NAIR; 00-CTEC
00-ESCPROJ	00-CTEC	00-NAIR; 00-ESCPROJ
00-NAIR	00-CTEC; 00-ESCPROJ	00-NAIR
01-ECOTOXICOLOGIA	00-CTEC; 00-ESCPROJ	00-NAIR; 01-ECOTOXICOLOGIA
02_03-DULCICOLA_INTEGRACAO	00-CTEC; 00-ESCPROJ	00-NAIR; 02_03-DULCICOLA_INTEGRACAO
03-MARINHO_ESTUARINO	00-CTEC; 00-ESCPROJ	00-NAIR; 03-MARINHO_ESTUARINO
04-PRAIAS	00-CTEC; 00-ESCPROJ	00-NAIR; 04-PRAIAS
05-MANGUEZAL	00-CTEC; 00-ESCPROJ	00-NAIR; 05-MANGUEZAL
05-RESTINGA	00-CTEC; 00-ESCPROJ	00-NAIR; 05-RESTINGA
06-MEGAFUNA	00-CTEC; 00-ESCPROJ	00-NAIR; 06-MEGAFUNA
07-ICTIOFAUNA_DULCICOLA	00-CTEC; 00-ESCPROJ	00-NAIR; 07-ICTIOFAUNA_DULCICOLA
07-ICTIOFAUNA_MARINHA	00-CTEC; 00-ESCPROJ	00-NAIR; 07-ICTIOFAUNA_MARINHA
08-SEDIMENTACAO_ABROLHOS	00-CTEC; 00-ESCPROJ	00-NAIR; 08-SEDIMENTACAO_ABROLHOS
98-DADOS_PRE_PMBA		Todos em Rede Rio Doce Mar
99-GERAL	Todos em Rede Rio Doce Mar	00-NAIR; 00-CTEC; 00-ESCPROJ
99-TEMP		Todos em Rede Rio Doce Mar

Fonte: autoria própria.

Todas as contas do projeto terão permissão de escrita abaixo da estrutura correspondente ao seu Anexo/Grupo de Trabalho. Apenas os próprios Anexos/Grupos de Trabalho terão permissão de escrita em seus respectivos folders de equipe. Os usuários do CTEC (00-CTEC) e ESCPRO (00-ESCPRO) poderão visualizar as informações de todos os folders de equipe. Os usuários do NAIR (00-NAIR) poderão visualizar e alterar as informações de todos os folders de equipe. Demais usuários de Anexos não poderão visualizar outros folders de equipe que não seja o seu próprio.

Também foram criados três outros folders de equipe, incluindo todos os usuários como membros:

- 98-DADOS_PRE_PMBA: Folder para armazenamento e publicação de todos os materiais, dados e documentações anteriores ao projeto (informações pré-desastre e pré-PMBA), com acesso de escrita para todos os usuários.
- 99-TEMP: Folder para compartilhamento de informações entre os usuários, com acesso de escrita para todos os usuários.
- 99-GERAL: Folder para publicação de informações gerais pela coordenação do projeto, com acesso de leitura para todos os usuários e escrita apenas para NAIR, CTEC e ESCPRO.

É muito importante manter todos os arquivos organizados, seguindo a estrutura de pastas sugerida sempre que possível. Visando padronizar, minimamente, as estruturas de organização dos arquivos entre os diversos Anexos/Grupos de Trabalho, também foram criadas as subpastas abaixo das pastas de equipe. As subpastas criadas têm como finalidade dividir os dados gerados por cada subgrupo ou subprojeto do Anexo/Projeto/Grupo de Trabalho. Para exemplificar, a sequência de figuras abaixo (Figura 4, Figura 5, Figura 6 e Figura 7) apresenta os níveis da estrutura de subpastas da pasta de equipe “04-PRAIAS”.

Figura 4: Estrutura interna de pastas (primeiro nível) do Anexo 04 (Praias).



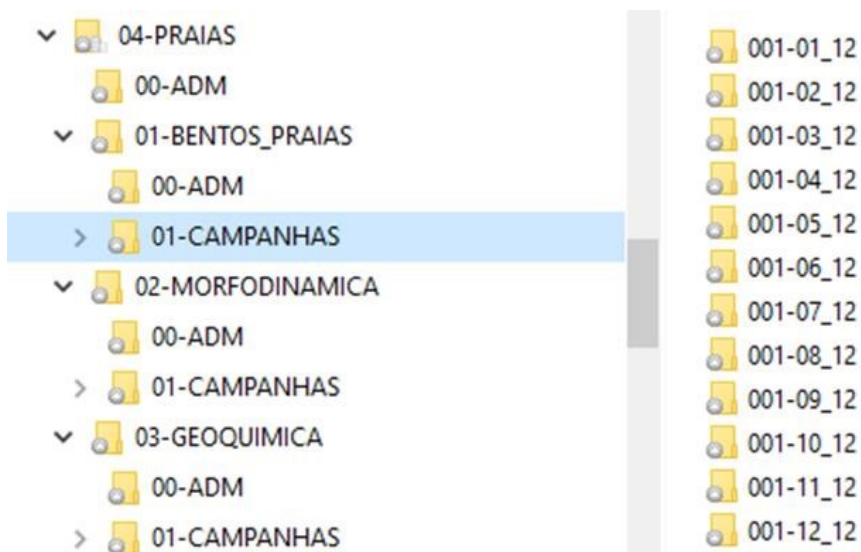
Fonte: autoria própria.

Figura 5: Estrutura interna de pastas (segundo nível) do Anexo 04 (Praias).



Fonte: autoria própria.

Figura 6: Estrutura interna de pastas (terceiro nível) do Anexo 04 (Praias).



Fonte: autoria própria.

Figura 7: Estrutura interna de pastas (quarto nível) do Anexo 04 (Praias).



Fonte: autoria própria.

Esta mesma estrutura foi replicada para cada um dos Anexos/Projetos. As subpastas “00-ADM” devem ser utilizadas para armazenamento de arquivos Administrativos e de Gestão, como relatórios, atas de reuniões, status report, etc., do Anexo/Grupo de Trabalho ou dos subgrupos.

As subpastas “01-CAMPANHAS” devem ser utilizadas para armazenamento de arquivos relativos às campanhas de monitoramento, como dados de coleta ou de análises geradas. Abaixo da estrutura de “01-CAMPANHAS” também foi sugerida a criação de subpastas no formato “<nº ano monitoramento>-<nº campanha>_<total campanhas ano>”. Abaixo desta estrutura foram criadas as subpastas para organização de dados brutos, análise de dados, fotos e vídeos relacionados a campanha de monitoramento.

Abaixo das pastas de equipe, também podem ser criadas outras pastas. Esta estrutura de pastas é apenas sugerida e pode ser alterada pelos usuários de forma a melhor atender aos objetivos e necessidades de cada Anexo/Grupo de Trabalho. Porém, é importante buscar, sempre que possível, manter a estrutura já criada e seguir o padrão de organização estabelecido durante a criação de novas pastas, visando facilitar a navegação de outros usuários externos ao seu Anexo/Grupo de Trabalho. Recomendamos que a estrutura de subdiretórios deve ter convenções de nomenclatura claras e documentadas. Arquivos ou diretórios separados podem se aplicar, por exemplo, a cada execução de um experimento, a cada versão de um conjunto de dados e/ou a cada pessoa no grupo.

- É importante reservar a extensão de arquivo de 3 letras para o formato de arquivo, como .txt, .pdf ou .csv
- Sempre identificar a atividade ou projeto no nome do arquivo.
- Identificar versões separadas de arquivos e conjuntos de dados usando convenções de nomenclatura de arquivos ou diretórios. Pode tornar-se rapidamente difícil identificar a versão "correta" de um arquivo.
- Registrar todas as alterações em um arquivo, não importando quão pequeno seja. Descartar versões obsoletas apenas após a realização de backups.

Maiores detalhes sobre a organização de usuários e de conteúdos no Dropbox pode ser consultado no anexo 26 (Guia de Utilização do Dropbox), na seção ANEXOS.

A organização dos data sets ou conjuntos de dados (Metadados e Dados) no Repositório de Dados também é crucial para garantir maior eficiência nas buscas e localização de conteúdo nesta camada de dados. Como esta organização de conteúdo envolve o gerenciamento de Metadados (dados sobre os dados), trataremos este assunto separadamente e de forma específica na próxima seção, DOCUMENTAÇÃO E METADADOS.

Garantia de Qualidade dos Dados:

No PMBA, deverão ser adotados processos adicionais, entre a coleta e publicação de dados, visando a garantia da qualidade dos dados e metadados. Estes processos incluem a captura e gravação padronizada de dados, e validação em 3 níveis da entrada de dados e da geração de metadados.

A seguir são apresentados os 4 passos entre a Coleta de Dados e a Publicação dos Dados e Metadados no Repositório de Dados, que garantem a padronização e validação do conteúdo.

- Processo de Coleta e Geração de Dados.

1. Geração de planilha de dados pré-formatada para o Anexo/Projeto, seguindo as orientações do CTEC e UGD. Abaixo, algumas das orientações:
 - a. Padronizar o cabeçalho seguindo os modelos disponibilizados pela CTEC e UGD.
 - b. Preencher todas as células de dados da planilha.
 - i. Manter apenas os campos chave necessários.
 - ii. Utilizar códigos de exceção para dados ausentes.
 - c. Informar domínio de valores e glossário de termos.
 - d. Não mesclar células de dados.
 - e. Evitar uma grande quantidade de colunas.
 - i. Dividir os dados em mais de uma aba se possível.
 - ii. Rever os campos chave de cada tabela.
2. Validação da planilha de dados pré-formatada após preenchimento pelo pesquisador.
 - a. Validação por outro(s) pesquisador(es) do mesmo Anexo/Projeto.
 - b. Validação pela equipe do CTEC.
 - c. Validação pela equipe do UGD.
3. Geração do Data Set (conjuntos de dados para o Repositório de Dados) com arquivos de dados e metadados seguindo os padrões do PMBA.
4. Validação do Data Set (arquivos de dados e metadados) após geração pelo pesquisador.
 - a. Validação por outro(s) pesquisador(es) do mesmo Anexo/Projeto.

- b. Validação pela equipe do CTEC.
- c. Validação pela equipe do UGD.

- Processo de Publicação dos Dados e Metadados no Repositório.

O processo de revisão acontece dentro e fora do Anexo/Projeto. Caso sejam detectados problemas em um dos 3 níveis de validação abaixo, o item de validação será retornado para o pesquisador responsável para ajustes e posteriormente o processo de validação é reiniciado.

1. Validação por outro(s) pesquisador(es) do mesmo Anexo/Projeto.
2. Validação pela equipe do CTEC.
3. Validação pela equipe do UGD.

No anexo 30 (Planilha de Dados Padrão), na seção **Erro! Fonte de referência não encontrada.**, é apresentado um exemplo de planilha pré-formatada para divulgação dos resultados acompanhada de todas as instruções para preenchimento. As regras de documentação e geração dos data sets (conjuntos de dados para o Repositório de Dados) com arquivos de dados e metadados relacionados serão abordadas na próxima seção, **Erro! Fonte de referência não encontrada.**

3.2 DOCUMENTAÇÃO E METADADOS

A documentação clara e detalhada é essencial para que os dados sejam compreendidos, interpretados e usados. A documentação de dados é comumente chamada de metadados, "dados sobre dados". A documentação de dados descreve em detalhes o conteúdo, os formatos e os relacionamentos internos e permite que outros pesquisadores encontrem, usem e citem os dados corretamente.

Os metadados, ou dados que descrevem dados, contêm informações sobre o conteúdo de um conjunto de dados (seu proprietário, administrador, extensão geográfica, unidades, etc.), bem como quem tem acesso aos dados (o proprietário, usuários selecionados ou o público alvo).

Nesta seção serão abordados os padrões de documentação e de metadados a serem utilizados no PMBA.

3.2.1 Quando se deve iniciar a documentação dos dados?

A documentação dos dados deve ser iniciada durante o planejamento da coleta e da geração dos dados, ou seja, ainda antes de se obter os dados. Isso torna o processo de documentação muito mais fácil. Para se construir a documentação no final do projeto, o processo será doloroso e detalhes importantes poderão ser perdidos ou esquecidos. Portanto não se deve esperar para documentar os dados.

3.2.2 Quais as ferramentas e padrões de documentação de metadados serão utilizados?

Conforme comentado anteriormente, o programa Morpho deverá ser utilizado para criar, editar e armazenar os metadados. O software Metacat será responsável pela publicação, busca, visualização e download online dos data sets.

O Morpho armazena os metadados em arquivos em conformidade com o padrão EML (Ecological Metadata Language), que é comumente utilizado para troca de informações relacionadas aos dados de monitoramento ambiental e de biodiversidade entre cientistas de todo o mundo. O EML é fruto de trabalhos prévios da Ecological Society of America. Maiores informações sobre o padrão EML podem ser consultadas no endereço abaixo:

<http://knb.ecoinformatics.org/software/eml/>

3.2.3 Que documentação e metadados acompanharão os dados?

A seguir são apresentadas as principais informações que deverão ser documentadas como metadados no PMBA. Trata-se do mínimo necessário para criar um conjunto de dados no Repositório de Dados.

Serão apresentadas algumas imagens com a interface do Metacat Data Catalog, módulo do Metacat disponível para documentação dos dados. A ferramenta Morpho possui a mesma estrutura e sequência para geração dos metadados, mas não exige que o usuário tenha conexão ativa com a internet. As duas ferramentas estarão disponíveis para cadastramento de metadados no PMBA.

Mas, é importante ressaltar que, mesmo na ausência de uma ferramenta para apoiar a geração dos metadados (Morpho ou Metacat), eles podem ser gerados, por exemplo, com a criação de arquivos texto "readme.txt", ou o equivalente, a serem disponibilizados em conjunto com os arquivos de dados.

- Dados Gerais:

A seguir são apresentadas as informações básicas ou gerais a serem documentadas como metadados (Figura 8).

- Título: Nome do conjunto de dados ou projeto de pesquisa que o produziu.

- Resumo: Descrição geral que sumariza a especificação dos conteúdos e propósito do conjunto de dados.

- Palavras-chave: Termos que ajudem na identificação do conjunto de dados. Devem ser utilizados, preferencialmente, termos que não estão contidos no Título.

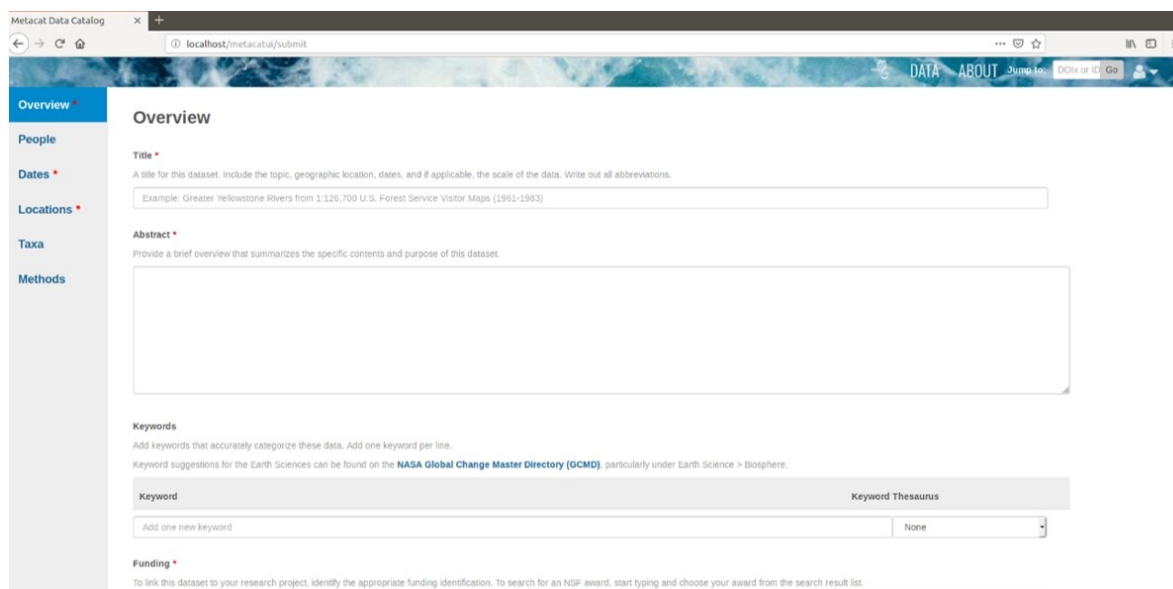
- Financiamento: Identificação da fonte de financiamento que possibilitou a coleta de dados.

- Data de publicação: Data de citação do conjunto de dados.

- Direitos de uso: Definição de como será usado e compartilhado o conjunto de dados.

- Identificador alternativo: Se disponível, informar identificador utilizado para encontrar informações relacionadas ao conjunto de dados em outras bases de dados.

Figura 8: Tela do Metacat para cadastro de informações gerais do data set.



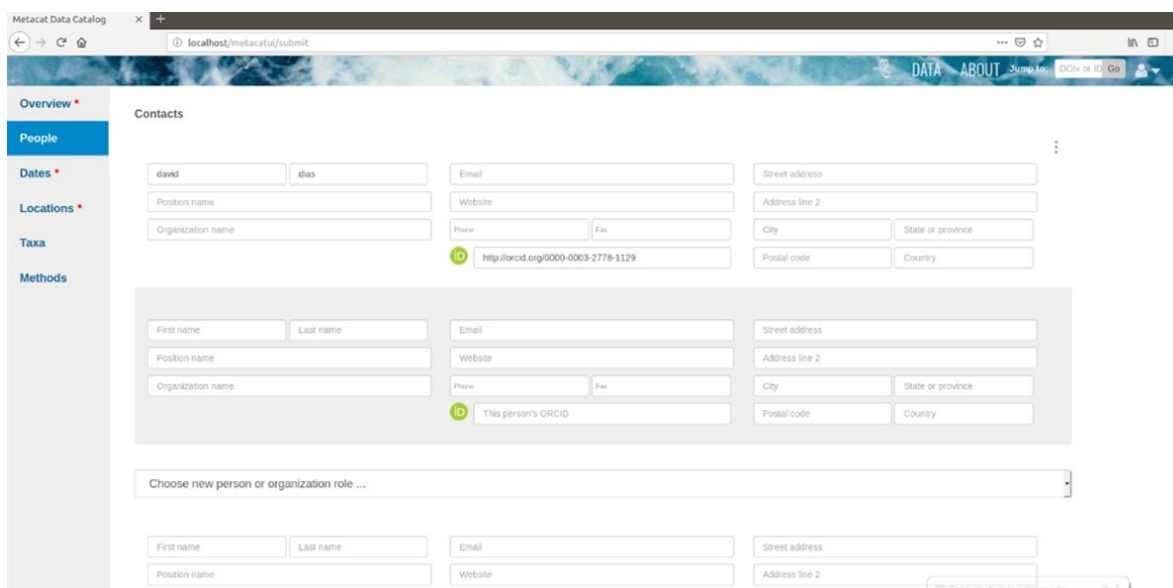
Fonte: autoria própria.

Dados de Pessoas:

A seguir são apresentadas as informações sobre as pessoas relacionadas ao conjunto de dados (Figura 9).

- Criadores, autores, contatos, etc.: Nomes e endereços das organizações ou pessoas que criaram os dados. Não existe um número limite de pessoas e não deve ser deixado de adicionar colaboradores.

Figura 9: Tela do Metacat para cadastro das pessoas associadas ao dataset.



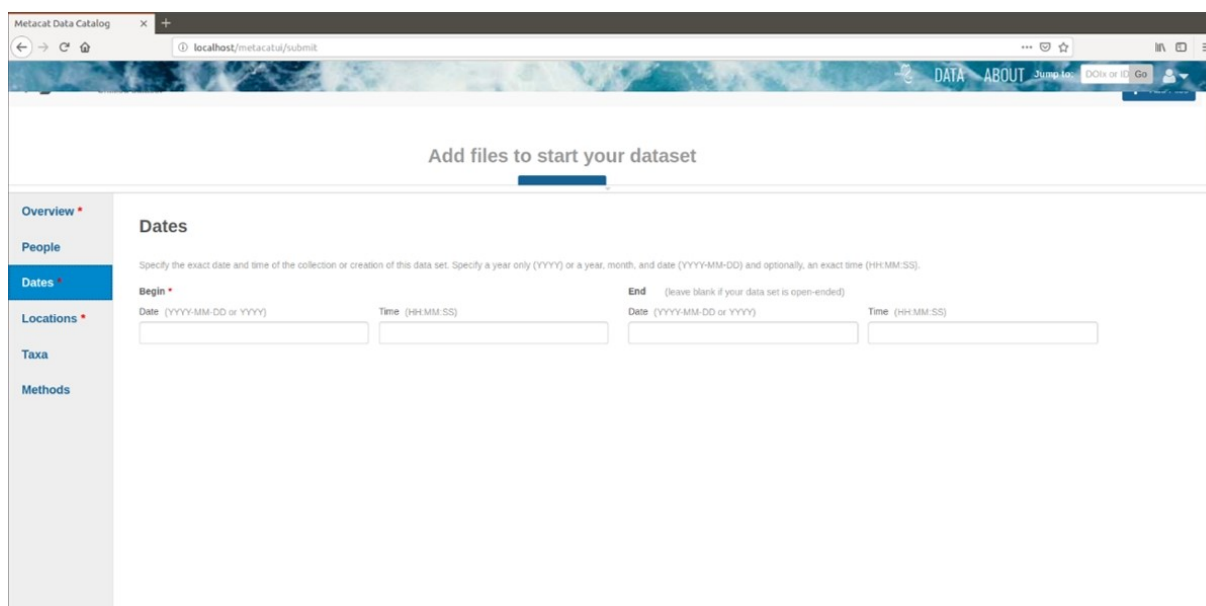
Fonte: autoria própria.

Dados de Cobertura Temporal:

A seguir são apresentadas as informações sobre o período de monitoramento (Figura 10).

- Datas: Datas-chave associadas aos dados, incluindo: data de início e término do projeto; data de lançamento; período coberto pelos dados; e outras datas associadas à vida útil dos dados, como ciclo de manutenção, agendamento de atualização. O formato preferido é “aaaa-mm-dd” ou “aaaa.mm.dd-aaaa.mm.dd” para um intervalo.

Figura 10: Tela do Metacat para cadastro da cobertura temporal do dataset.



The screenshot shows a web browser window with the URL localhost/metacatal/submit. The page title is 'Metacat Data Catalog'. The main heading is 'Add files to start your dataset'. On the left, there is a sidebar with navigation links: Overview, People, Dates (selected), Locations, Taxa, and Methods. The main content area is titled 'Dates' and contains the following text: 'Specify the exact date and time of the collection or creation of this data set. Specify a year only (YYYY) or a year, month, and date (YYYY-MM-DD) and optionally, an exact time (HH:MM:SS)'. Below this, there are two sections: 'Begin' and 'End'. Each section has a 'Date' field (format: YYYY-MM-DD or YYYY) and a 'Time' field (format: HH:MM:SS). The 'End' section has a note: '(leave blank if your data set is open-ended)'. The form fields are currently empty.

Fonte: autoria própria.

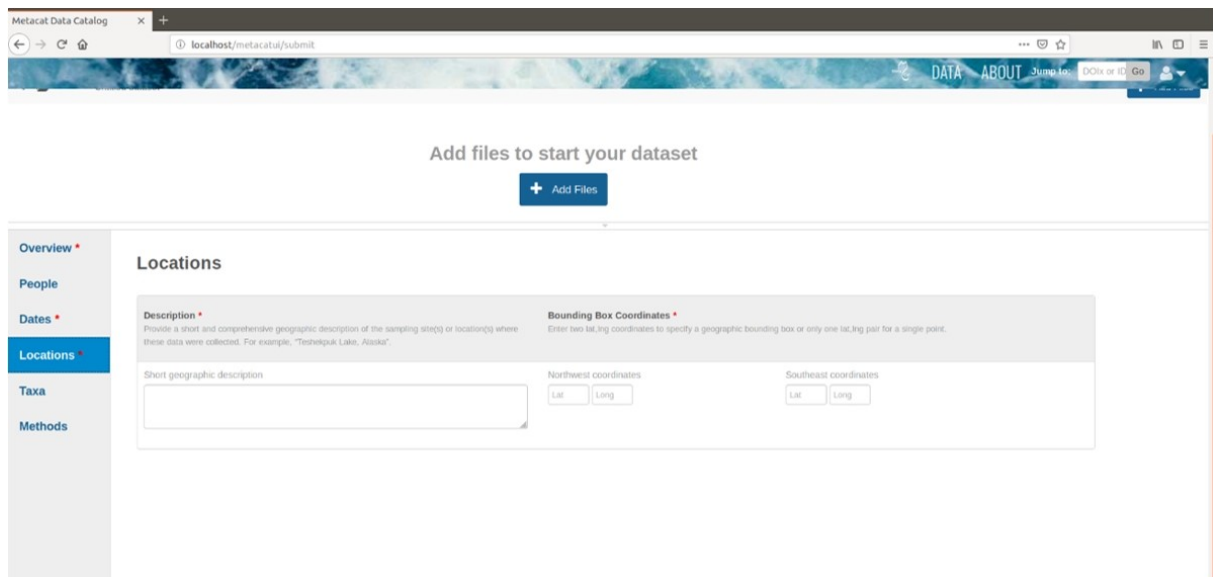
Dados de Cobertura Geográfica:

A seguir são apresentadas as informações sobre os pontos ou regiões de monitoramento (Figura 11).

- Descrição: Texto resumido descrevendo os locais onde houve coleta dos dados.

- Pontos e/ou caixa de coordenadas: Latitude e longitude dos locais de coleta.

Figura 11: Tela do Metacat para cadastro da cobertura geográfica do dataset.



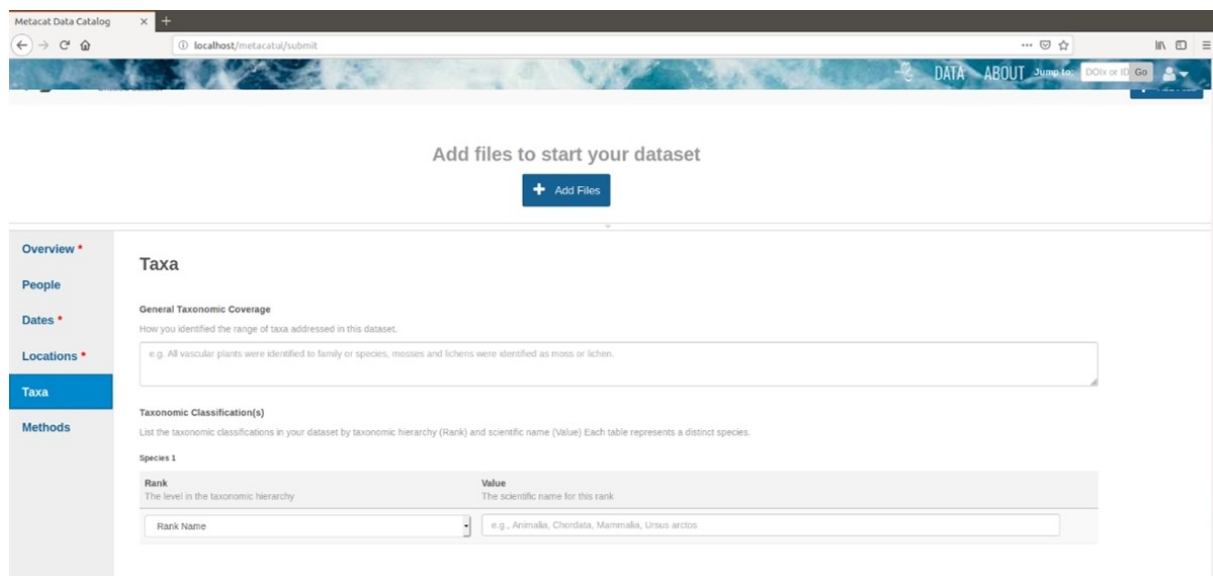
Fonte: autoria própria.

Dados de Cobertura Taxonômica:

A seguir são apresentadas as informações sobre as espécies monitoradas (Figura 12).

- Resumo: Descrição de como foi realizada a identificação dos táxons.
- Classificação: lista contendo a classificação taxonômica com os campos de rank e valor.

Figura 12: Tela do Metacat para cadastro da cobertura taxonômica do dataset.



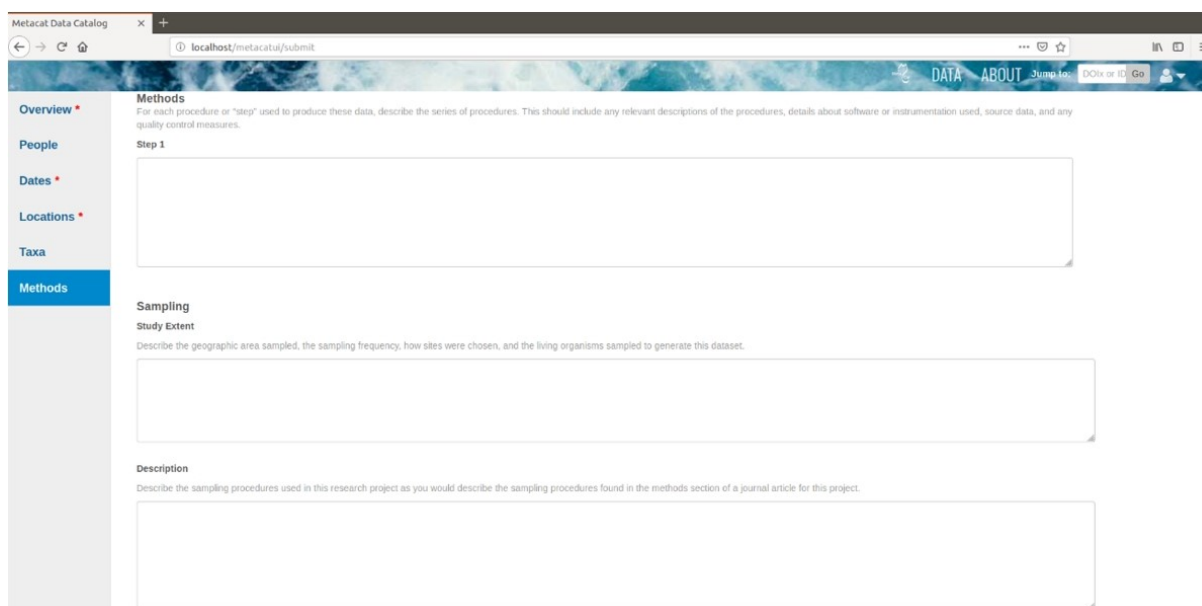
Fonte: autoria própria.

Dados de Metodologias:

A seguir são apresentadas as informações sobre as metodologias utilizadas para monitoramento e geração dos dados (Figura 13).

- Etapas: Descrição das etapas utilizadas para a produção ou coleta dos dados.
- Extensão do estudo: Descrição da área geográfica amostrada, frequência de amostragem, como os locais são escolhidos e os organismos amostrados para gerar o conjunto de dados.
- Descrição: Resumo dos procedimentos de amostragem utilizados no processo de pesquisa para produção do conjunto de dados.

Figura 13: Tela do Metacat Data Catalog para cadastro das metodologias utilizadas



Fonte: autoria própria.

3.2.4 Como identificar unicamente cada conjunto de dados?

Para que seja possível compartilhar ou citar um conjunto de dados, é necessário atribuir um identificador exclusivo, persistente e público a ele. Há uma variedade de esquemas de identificadores públicos, mas propriedades comuns de bons esquemas são:

- Acionáveis (você pode "clique" neles em um navegador da Web).
- Globalmente exclusivo em toda a Internet.
- Persistente por pelo menos a vida dos seus dados.

A Unidade de Gestão de Dados (UGD) recomenda a utilização do DOI (Digital Object Identifier). Trata-se de um identificador que se torna acionável quando incorporado em uma URL ou link Web. Os DOIs são muito populares na publicação de periódicos acadêmicos. Eles são resolvidos em "http://dx.doi.org". A resolução depende do redirecionamento de HTTP e do identificador de identificador e pode ser gerenciada por meio de uma API ou interface de usuário.

3.3 ÉTICA E DIREITOS AUTORAIS

Nesta seção serão abordadas as questões éticas e de direitos autorais relacionadas aos dados do PMBA.

3.3.1 Como serão administradas as questões éticas?

Os investigadores que realizam pesquisas envolvendo participantes humanos devem solicitar o consentimento para preservar e compartilhar os dados. Questões éticas podem afetar como os dados são armazenados e transferidos, como são acessados, como são utilizados e por quanto tempo deverão ser preservados.

O gerenciamento de preocupações éticas pode incluir: anonimato de dados; encaminhamento a comitês de ética departamental ou institucional; e acordos formais de consentimento.

Vale ressaltar que todos os dados brutos coletados pela RRDM no âmbito do Programa de Monitoramento são encaminhados, a cada 3 meses, em paralelo, para Fundação Renova, CIF (Comitê Interfederativo) e CT-Bio (Câmara Técnica de Conservação e Biodiversidade). Todo o material que a RRDM encaminha à Fundação Renova e ao CIF/CT-Bio, quando aprovado, se torna público. Antes dos dados se tornarem públicos, caso seja identificada a necessidade ou interesse em compartilhamento e/ou divulgação de informações para sociedade, instituições e/ou pesquisadores externos ao projeto, deve ser solicitada autorização prévia ao CIF/CT-Bio, por intermédio da Fundação Renova ou diretamente. O conteúdo das publicações não passa por censura da Fundação Renova.

3.3.2 Como serão gerenciados os direitos autorais?

Os dados devem ser licenciados. É comum que os provedores de dados apliquem licenças para limitar como os dados poderão ser utilizados e proteger a privacidade dos participantes do estudo ou para orientar os usos posteriores sobre os dados (por exemplo, exigir atribuição ou proibir fins lucrativos).

A seguir serão apresentadas algumas opções de licenciamento para gestão dos direitos autorais. É importante que o PMBA defina, em comum acordo com Fundação Renova, CIF/CT-Bio e RRDM, qual padrão de licenciamento será utilizado para os dados gerados. Esta definição poderá ser realizada após a finalização do projeto, mas é importante que ocorra antes da publicação e divulgação dos dados a sociedade em geral.

Licenças Creative Commons:

Atualmente é possível escolher entre seis licenças da Creative Commons. São licenças que permitem uso, modificação e distribuição dadas as condições específicas impostas por cada uma. Abaixo são descritas as condições para cada licença.

- CC BY: é necessária a citação dos autores sempre que os dados forem utilizados. Exceção caso os autores permitam explicitamente o uso sem citação.

- CC BY-SA: além dos requerimentos da licença anterior é necessário que a distribuição de qualquer modificação seja feita nos mesmos termos.
- CC BY-ND: além dos requerimentos de citação da primeira licença, é proibida a modificação dos dados originais. Exceção caso os autores permitam explicitamente.
- CC BY-NC: requerimentos da licença CC BY com exclusão de uso comercial sem permissão explícita.
- CC BY-NC-SA: requerimentos da licença anterior com obrigação da utilização dos mesmos termos.
- CC BY-NC-ND: requerimentos da licença CC BY-NC com exclusão de modificações dos dados originais.

A utilização das licenças Creative Commons facilita a disseminação dos dados e inclui um arcabouço legal mais forte do que as licenças criadas pelos publicadores. Instituições como GBIF (Global Biodiversity Information Facility) e DataONE começaram a endossar o uso dessas licenças por falta de uma escolha dos usuários e na facilidade de interpretação das mesmas.

No anexo 36 (Licenças Creative Commons), na seção ANEXOS, são apresentados maiores detalhes sobre cada uma das opções de licenciamento fornecidas pela Creative Commons.

Licença Brasileira do Programa de Pesquisa em Biodiversidade (PPBio):

O PPBio possui uma política de dados publicada oficialmente no diário oficial da união que descreve como os dados coletados devem ser utilizados e disponibilizados. A publicação dessa política de dados é um exemplo de construção própria de licenciamento para um repositório de dados.

Qualquer uso não educacional deve ser acordado entre as partes (interessado e autores do conjunto de dados). Evidencia a proteção do investimento público e propriedade intelectual dos autores. Não existe a obrigação de dar permissão para uso comercial.

É respeitado o período de embargo de dados sensíveis, restringindo acesso aos mesmos e liberação mediante autorização do Comitê Gestor de Informação (CGI). Dados sensíveis são definidos como aqueles em que sua divulgação pública pode causar efeito adverso ao local e/ou às comunidades de origem das mesmas.

O acesso aos dados por meio da internet implica no aceite das condições de uso e termo de compromisso. Essa é uma condição comum em quase todos os portais na internet. Caso o usuário não concorde com alguma das cláusulas, deve deixar o portal imediatamente.

É excluída a responsabilidade de qualquer dano, consequência ou prejuízo na utilização dos dados. Cláusula comumente encontrada em ferramentas/portais que disponibilizam de forma aberta seus dados. O objetivo é não causar a inibição dos possíveis publicadores por meio de represarias.

Atribuição de créditos é expressamente obrigatória e não negociável. Outra cláusula para proteção do investimento público e da propriedade intelectual. É recomendado também aos interessados na utilização dos dados informar os autores que seu dado foi utilizado como forma de reconhecimento.

A definição de prazos é o ponto mais restrito da política de dados e foi criada com o objetivo de evitar longas esperas para a disponibilização dos dados. Os metadados devem ser disponibilizados para publicação on-line no prazo máximo de 30 dias após a coleta dos dados e disponibilizados via internet em no máximo 7 dias após a entrega para publicação ao CGI. Prazo de 12 meses para dados preliminares serem disponibilizados ao CGI e 24 meses para dados consolidados. Com exceções caso o Conselho Diretor do PPBio assim o faça.

No anexo 37 (Política de Dados do PPBio), na seção ANEXOS, são apresentados maiores detalhes sobre política de dados do Programa de Pesquisa em Biodiversidade (PPBio) publicada no Diário Oficial da União.

Licença Brasileira do Programa de Pesquisas Ecológicas de Longa Duração (PELD):

Política de dados definida por Resolução Normativa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) visando a disponibilização dos dados provenientes do projeto PELD pela sua importância na gestão pública e científica.

Os metadados devem ser disponibilizados até metade da vigência original do projeto ou por solicitação do CNPq. Dados devem ser disponibilizados em até 60 dias da vigência final do projeto. Os dados preliminares permanecem privados durante o período de embargo de até dois anos após o término da vigência do projeto e prorrogável por mais um ano.

Dados sensíveis são privados e necessitam de pedido de acesso explícito dos autores e gestor do sítio. Cabe ressaltar que as restrições de acesso podem ser a nível de registro, normalmente o acesso é especificado a nível de planilha ou o conjunto metadados e dados.

A utilização dos dados para fins comerciais não é permitida sem autorização explícita dos autores. Uso não autorizado sujeito a sanções legais.

Os dados são cedidos somente caso o usuário realize um cadastro e a distribuição a terceiros não é permitida. Essa cláusula cria um atrito a disseminação do dado. Atualmente, poucos repositórios conseguem atender a essa demanda tecnicamente.

O reconhecimento da autoria é obrigatório e também deve incluir uma citação do repositório de origem. É recomendado o convite dos autores a participar intelectualmente dos trabalhos desenvolvidos a partir dos dados utilizados.

Ao utilizar os dados o usuário concorda com os termos do mesmo, compromete-se a notificar os autores em caso de publicação. O repositório também vai informar os autores toda vez que ocorrer um download dos seus dados.

No anexo 38 (Política de Dados do PELD), na seção ANEXOS, são apresentados maiores detalhes sobre política de dados do Programa de Pesquisa Ecológica de Longa Duração (PELD) definida por Resolução Normativa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

3.4 SELEÇÃO E PRESERVAÇÃO

Além de quaisquer requisitos impostos externamente, é necessário avaliar a utilidade a longo prazo dos dados. Por exemplo, se os dados forem provenientes de um experimento que seja reproduzível com maior rapidez, baixo custo e precisão conforme a tecnologia avança, é possível armazená-lo por um período relativamente breve. Já, se os dados consistirem em observações feitas fora do laboratório que nunca poderão ser repetidas, deve-se avaliar o seu armazenamento indefinidamente. Nesta seção serão abordadas questões relacionadas a seleção e preservação dos dados, como o tipo do dado a ser preservado, local de preservação, tempo e formato de retenção e equipes responsáveis.

3.4.1 Quais dados são de valor a longo prazo e devem ser mantidos, compartilhados e/ou preservados?

As regras de preservação de dados do projeto foram especificadas de acordo com as camadas de dados da arquitetura. E estão relacionadas ao valor potencial de reutilização dos dados ao longo do projeto e a sua disponibilidade para pesquisas e análises mesmo após o projeto.

Armazém de Dados:

Como já apresentado anteriormente, esta é a primeira camada de gestão de dados, onde os dados disponibilizados estarão em seu formato mais bruto, com poucos ou nenhum tratamento e sem os metadados relacionados (informações sobre os dados).

Esta camada será muito utilizada pelos pesquisadores durante os 5 anos de monitoramento do PMBA, armazenando dados em formatos de arquivos que facilitem a geração dos relatórios periódicos e dos dados finais para publicação no Repositório de Dados, local oficial de publicação dos resultados do monitoramento.

Grande parte dos arquivos desta camada são temporários e necessários apenas enquanto os dados finais ainda estão sendo trabalhados para publicação. Após a publicação dos relatórios e resultados no Repositório de Dados, a necessidade por esses dados diminui.

Portanto, a UGD entende que os dados da camada deverão estar disponíveis para acesso durante todo o período do PMBA (5 anos). Após este período, deverá ser realizado um backup dos dados para preservação por tempo indeterminado. Este backup já não estará mais disponível para acesso pelos pesquisadores via ferramenta Dropbox. Caso seja necessário acessar os dados do backup, deverá ser solicitada a recuperação do conteúdo à equipe responsável pela gestão de dados.

Repositório de Dados:

Como visto, são áreas de armazenamento da produção científica do PMBA, armazenando metadados e dados de diferentes formatos que possam ser reaproveitados por cientistas, profissionais envolvidos com o manejo ambiental e formuladores de políticas públicas de dentro e de fora do projeto.

O Repositório de Dados já é o resultado da seleção de conteúdo da camada anterior (Armazém de Dados) com uma série de processos para enriquecimento dos dados, como a adição de metadados, validações de conteúdo, padronizações e tratamento de qualidade.

Por considerar que os dados de resultado do monitoramento poderão ser utilizados para muito além do período do PMBA, o Repositório de Dados deverá ser preservado por tempo indeterminado. Serão realizados backups periódicos da camada apenas por segurança, mas todos os dados permanecerão acessíveis pela ferramenta Metacat e Morpho por tempo indeterminado.

O Repositório de Dados proporciona visibilidade aos resultados de pesquisas e possibilita a preservação da memória científica de sua instituição ou projeto.

Banco de Dados:

As bases de dados relacionais permitem que diversas aplicações tenham acesso aos dados possibilitando que sejam utilizadas durante décadas ou longos períodos. O Banco de Dados, diferentemente das camadas anteriores, deve ser utilizado principalmente para a realização de análises de dados utilizando ferramentas analíticas.

De forma similar ao Repositório de Dados, a camada também é resultado da seleção de conteúdo de camadas anteriores, com a vantagem de agregar relacionamentos pré-definidos entre as informações facilitando a geração de análises integradas entre assuntos.

Análises integradas de dados poderão ser realizadas ao longo do projeto e após a sua finalização. Por isso, os dados desta camada também deverão estar disponíveis por tempo indeterminado. Serão realizados backups periódicos da camada apenas por segurança, mas todos os dados permanecerão acessíveis pelo Tableau ou por outras ferramentas analíticas.

3.4.2 Qual é o plano de preservação a longo prazo do conjunto de dados (dataset)?

Os dados das camadas de Repositório de Dados e Banco de Dados deverão ser preservados em seu estado original por tempo indeterminado. Desta forma, todas as aplicações utilizadas para acessar e consumir as informações geradas estarão disponíveis ao longo do projeto e após o projeto. Os custos de manutenção e suporte desta infraestrutura, incluindo backups regulares, serão detalhados na seção CUSTOS DE IMPLEMENTAÇÃO.

Com relação ao Armazém de Dados, após os 5 anos de projeto, toda a infraestrutura poderá ser desativada e a licença do Dropbox não mais renovada. Se isto acontecer, deverá ser realizado um backup e compactação de todos os dados disponíveis no ambiente antes da sua desativação. Este

backup será preservado (armazenado) por tempo indeterminado. Os custos para implementação desta política de preservação, além dos custos de recuperação do conteúdo, quando necessário, também serão detalhados na seção CUSTOS DE IMPLEMENTAÇÃO.

Os procedimentos de armazenamento e backup serão detalhados na seção seguinte ARMAZENAMENTO E BACKUP.

3.5 ARMAZENAMENTO E BACKUP

Armazenar dados em laptops, discos rígidos de computador ou dispositivos de armazenamento externos é muito arriscado. O uso de armazenamento robusto e gerenciado com backups regulares são essenciais para garantir a disponibilidade dos dados sem interrupções, dando mais segurança aos usuários das informações. Nesta seção serão abordadas questões relacionadas ao armazenamento seguro de dados, políticas de backups e gestão de acessos.

3.5.1 Como os dados serão armazenados durante a pesquisa?

O Quadro 8 apresenta um resumo das áreas de armazenamento definidas para o PMBA e como os volumes estarão distribuídos entre elas. Também apresenta as estimativas de volumetria e as áreas em disco já contratadas ou previstas para contratação.

Quadro 8: Áreas de armazenamento e de backup estimadas para as camadas.

Camada de Dados	Ferramentas	Área Útil Estimada	Área Útil Contratada ou a Contratar	Área de Backup
Armazém de Dados	Dropbox	160 TBytes	Ilimitada	320 TBytes
Repositório de Dados	Morpho; Metacat	10 TBytes	20 TBytes	20 TBytes
Banco de Dados	PostgreSQL	10 TBytes	20 TBytes	20 TBytes

Fonte: autoria própria.

Armazém de Dados:

O Dropbox já está contratado para o PMBA e conta com espaço ilimitado para os dados. Como foi estimado um total de 160 TBytes de dados na camada para os 5 anos de projeto, será necessária uma área de backup e 320 TBytes, permitindo a manutenção de duas áreas distintas, sendo uma no servidor de arquivos na nuvem e outra no servidor de dados local.

Maiores detalhes sobre a organização de usuários e de conteúdos no Dropbox podem ser consultados na seção 3.1.2, item Organização dos Dados, e no anexo 26 (Guia de Utilização do Dropbox), na seção ANEXOS.

A Unidade de Gestão de Dados (UGD) recomenda ainda a utilização de ferramentas de compactação de arquivos (por exemplo, ZIP, GZIP, TAR) para conjuntos de dados que não serão acessados com

frequência, como forma de economizar espaço em disco e facilitar as transferências e compartilhamentos (downloads e uploads mais rápidos).

Repositório de Dados:

O Repositório de Dados ainda está em fase de implementação pela equipe da UGD, mas estima-se a necessidade de 10 TBytes de dados na camada. Como já comentado, é recomendado a implementação de dois servidores Metacat, sendo um deles local e outro na nuvem, com sincronização de dados entre eles. Assim, será necessária uma área útil de 20 TBytes para armazenamento. Além deste volume, serão necessários outros 20 TBytes para manutenção de duas áreas de backups distintas, sendo uma no servidor de arquivos na nuvem e outra no servidor de dados local.

Maiores detalhes sobre a organização dos data sets ou conjuntos de dados (Metadados e Dados) no Repositório de Dados pode ser consultado na seção DOCUMENTAÇÃO E METADADOS.

Banco de Dados:

O Banco de Dados também está em fase de implementação pela equipe da UGD. Assim como o Repositório de Dados, estima-se a necessidade de 10 TBytes de dados na camada e também serão implementados dois servidores, sendo um deles local e outro na nuvem, com sincronização de dados entre eles, o que exigirá uma área útil de 20 TBytes para armazenamento e outros 20 TBytes para manutenção das duas áreas de backups (nuvem e local).

A organização de conteúdo na camada de Banco de Dados será realizada durante a etapa de Modelagem de Dados, que será iniciada posteriormente a entrega deste Plano de Gestão de Dados.

3.5.2 Como serão realizados os backups de dados durante a pesquisa?

Realizar backups regulares é parte integrante do gerenciamento de dados. É possível criar backups de dados utilizando computadores pessoais, discos rígidos externos ou servidores de dados. E recomenda-se a utilização de serviços de backup automáticos fornecidos pelos próprios sistemas operacionais, sistemas de repositórios de dados e sistemas de bancos de dados.

A boa prática é ter sempre duas cópias (backups) em pelo menos dois locais distintos. Distribuindo geograficamente os backups, por exemplo com cópias locais e remotas, se reduz o risco de calamidade em um mesmo local (falta de energia, inundação, incêndio, etc.).

Para ter certeza de que o sistema de backup está funcionando, também é necessário recuperar periodicamente seus arquivos de dados para confirmar a integridade e viabilidade de leitura.

No PMBA, a Unidade de Gestão de Dados (UGD) será responsável pela realização dos backups e restaurações de dados. E sempre serão utilizados os servidores de dados a serem contratados para o projeto, sendo um deles na nuvem e outro local. Os backups deverão ser realizados semanalmente e sempre seguidos de um teste de recuperação. Para a realização dos backups, deverão ser utilizadas

funcionalidades nativas já disponíveis nos próprios softwares de gestão das camadas de dados (Dropbox, Metacat, PostgreSQL). Abaixo, um resumo das regras de backup adotadas.

- Quantidade de cópias de segurança (backups): 2 cópias por camada de dados.
- Localização: 1 cópia em servidor de dados na nuvem e 1 cópia em servidor de dados local.
- Periodicidade de backup: Semanal.
- Periodicidade de teste de restauração: Semanal, a cada novo backup.
- Ferramentas utilizadas: Funcionalidades nativas do Dropbox, Metacat e PostgreSQL.
- Equipe responsável: Unidade de Gestão de Dados (UGD).

Vale ressaltar que a solução contratada para o Armazém de Dados (Dropbox) possui uma funcionalidade de versionamento dos dados, o que reduz a necessidade de backups regulares. O versionamento do Dropbox gera versões de cada arquivo armazenado na ferramenta sempre que uma alteração é realizada em seu conteúdo. E permite a recuperação de qualquer versão histórica dos arquivos em caso de incidentes. Além disso, a ferramenta será utilizada para garantir a sincronização dos backups do projeto entre servidor local e servidor na nuvem.

Com a integração futura do Repositório de Dados do PMBA ao servidor central do DataONE na Rede Internacional de Biodiversidade, será possível contar com um nível de segurança ainda maior para os dados, pois a solução garantirá a sincronização dos dados com outros nós da rede internacional, ou seja, teremos cópias de segurança (backups) de cada metadado em locais físicos diferentes do globo.

3.5.3 Como serão gerenciados os acessos e a segurança?

A segurança de dados é a proteção dos dados contra acesso, uso, alteração, divulgação e destruição não autorizados.

É necessário garantir e se certificar de que os dados estejam seguros em relação aos critérios abaixo.

Segurança de Rede:

É de extrema importância manter dados confidenciais protegidos por contas de usuários e senhas, com funcionalidade de rastreabilidade e auditoria para identificação proativa de possíveis falhas de segurança. Em casos extremos, materiais confidenciais devem ser disponibilizados em computadores desconectados da internet.

Todas as soluções de gestão de dados adotadas pelo PMBA (ALLIMS, Dropbox, Morpho, Metacat, PostgreSQL, Tableau), sem exceção, só podem ser acessadas através de credenciais de acesso com usuário e senha, controlados por administradores pertencentes a Unidade de Gestão de Dados (UGD) e com funcionalidade de rastreabilidade e auditoria habilitadas.

Além disso, até que os dados do Repositório de Dados e do Banco de Dados sejam declarados de domínio público pela Fundação Renova, tais camadas não estarão disponíveis à sociedade pela internet. As camadas estarão acessíveis apenas às redes locais de Universidades e demais instituições integrantes da Rede Rio Doce Mar (RRDM).

Segurança Física:

O acesso a prédios e salas onde servidores, computadores e mídias de dados estejam armazenados devem ser restringidos. Além disso, apenas pessoas confiáveis e autorizadas devem atuar na resolução de problemas e manutenções nos servidores de dados.

Os servidores de dados do projeto deverão estar centralizados e localizados em dependências com acesso apenas a colaboradores do ESCPRO, CTEC e NAIR. Além disso, e justamente por se preocupar com a segurança física de seus servidores e dados, que a Unidade de Gestão de Dados (UGD) recomenda a utilização de uma réplica do servidor local na nuvem. Os centros de processamento de dados disponibilizados em nuvem contam com políticas de segurança física rigorosas e homologados internacionalmente.

Sistemas e Arquivos de Dados:

A seguir, práticas seguidas pela equipe da UGD para garantir proteção aos arquivos de dados em todas as camadas.

- Manter a proteção contra vírus sempre atualizada em todos os servidores de dados.
- Manter os sistemas e softwares de servidores sempre atualizados.
- Não enviar nem compartilhar dados confidenciais por pen drives, e-mail, Skype, Whatsapp e aplicações similares. O compartilhamento de dados deve ser sempre realizado por ferramentas especializadas, como o Dropbox, que protegem os dados com contas de usuários e senhas e possuem a funcionalidade de rastreabilidade habilitada.

A UGD orienta que as recomendações acima também devem ser aplicadas a todos os computadores pessoais e por todos os integrantes do PMBA.

3.6 COMPARTILHAMENTO DE DADOS

São vários os motivos por que se deve compartilhar os dados gerados por projetos de pesquisa.

- Atendimento a requerimentos dos editores e financiadores.
- Atendimento a requerimentos de agências e instituições governamentais.
- Disponibilização para que os dados sejam reutilizados por outras pesquisas e projetos, visando responder a novos questionamentos científicos.

- Divulgação na comunidade científica para citações e reconhecimento por outros pesquisadores.

Nesta seção serão abordadas questões relacionadas ao compartilhamento de dados, incluindo discussões sobre as formas de compartilhamento e suas restrições, além de como identificar e dar crédito aos autores dos conjuntos de dados (citação de dados).

3.6.1 Como os dados serão compartilhados?

Os métodos utilizados para compartilhar dados dependerão de vários fatores, como tipo, tamanho, complexidade e sensibilidade dos dados. A seguir trataremos das várias formas de se compartilhar e acessar os dados no PMBA.

Formas de Compartilhamento:

É possível compartilhar dados de várias formas e utilizando diversas ferramentas computacionais.

- E-mail para solicitantes individuais.
- Publicações por meio de um projeto ou site pessoal.
- Publicações como material suplementar para ser hospedado no site de um editor de periódico.
- Compartilhamento de meios digitais físicos, como pen drives.
- Compartilhamento por ferramentas de comunicação digital, como Whatsapp, Skype, etc.
- Depósito e compartilhamento em um armazém de dados do projeto.
- Depósito e compartilhamento em um repositório de dados com metadados associados.

Enquanto as cinco primeiras opções acima são formas válidas de compartilhar dados, um armazém de dados como o Dropbox ou um repositório de dados como o Metacat são muito mais especializados em fornecer acesso seguro e de longo prazo. Os dados depositados nestas camadas de dados podem ser complementados com metadados, facilitando o consumo das informações por outros usuários, e além disso, protegem os dados com contas de usuários e senhas e possuem a funcionalidade de rastreabilidade habilitada.

Por isso, para o PMBA, apenas as duas últimas opções devem ser consideradas como formas de compartilhamento de dados. Não se deve enviar nem compartilhar dados confidenciais por pen drives, e-mail, Skype, Whatsapp e aplicações similares.

Camadas de Dados e Compartilhamento:

Existem algumas diferenças entre as camadas de dados com relação ao compartilhamento de conteúdo.

- Armazém de Dados: No Dropbox, ferramenta de armazém de dados contratada para o PMBA, os dados compartilhados ainda não possuem, necessariamente, metadados associados. Entende-se que

os dados ainda estão em sua forma inicial e ainda serão trabalhados e formatados para apresentação final dos resultados. Trata-se em sua grande parte de dados temporários e de trabalho, em constante alteração. Portanto, o compartilhamento dos dados desta camada ocorre principalmente internamente nas equipes de cada Anexo/Projeto, como forma de garantir maior colaboração para finalização dos resultados que serão futuramente publicados.

- Repositório de Dados: No Metacat, ferramenta de repositório de dados contratada para o PMBA, os dados compartilhados já possuem metadados associados (disponibilizados em conjunto com a ferramenta Morpho). Entende-se que os dados publicados já são resultado dos trabalhos de pesquisa de cada Anexo/Projeto e devem ser divulgados inclusive para equipes externas ao Anexo/Projeto de origem.

- Banco de Dados: O compartilhamento dos dados armazenados nesta camada será realizado por ferramentas analíticas (por exemplo o Tableau) que acessarão os dados para geração de análises e outras integrações. Após a utilização das ferramentas analíticas, os produtos gerados (análises de dados e relatórios) deverão ser publicados no Repositório de Dados para divulgação e compartilhamento com demais equipes e interessados.

Grupos de Acesso e Usuários Relacionados:

O PMBA possui um padrão para concessão de acessos aos dados e informações, baseado em grupos de usuário e usuários relacionados.

Foram definidos 14 grupos de usuários, como relacionados a seguir:

- 00-CTEC: Grupo que representa os usuários vinculados à CTEC.
- 00-ESCPROJ: Grupo que representa os usuários vinculados ao ESCPRO.
- 00-NAIR: Grupo que representa os usuários vinculados ao NAIR.
- 01-ECOTOXICOLOGIA: Grupo que representa os usuários vinculados ao Anexo/Projeto 1 (Ecotoxicologia).
- 02_03-DULCICOLA_INTEGRACAO: Grupo que representa os usuários vinculados ao Anexo/Projeto 3 (Dulcícola e Integração).
- 03-MARINHO_ESTUARINO: Grupo que representa os usuários vinculados ao Anexo/Projeto 3 (Marinho e Estuarino).
- 04-PRAIAS: Grupo que representa os usuários vinculados ao Anexo/Projeto 4 (Praias).
- 05-MANGUEZAL: Grupo que representa os usuários vinculados ao Anexo/Projeto 5 (Manguezal).
- 05-RESTINGA: Grupo que representa os usuários vinculados ao Anexo/Projeto 5 (Restinga).
- 06-MEGAFUNA: Grupo que representa os usuários vinculados ao Anexo/Projeto 6 (Megafauna).

- 07-ICTIOFAUNA_DULCICOLA: Grupo que representa os usuários vinculados ao Anexo/Projeto 7 (Dulcícola).
- 07-ICTIOFAUNA_MARINHA: Grupo que representa os usuários vinculados ao Anexo/Projeto 7 (Marinha).
- 08-SEDIMENTACAO_ABROLHOS: Grupo que representa os usuários vinculados ao Anexo/Projeto 8 (Sedimentação em Abrolhos).
- Todos em Rede Rio Doce Mar (RRDM): Grupo que representa todos os usuários da RRDM.

Todo usuário do PMBA deverá estar associado a pelo menos um dos grupos acima. E cada conteúdo gerado também deverá estar associado a pelo menos um dos grupos de acesso, com permissões apenas de leitura ou leitura e escrita (o que garante a realização de alterações pelos usuários daquele grupo).

Desta forma, é possível diferenciar os níveis de acesso (somente leitura; ou leitura e escrita) por Anexo/Projeto. Devido à grande quantidade de pesquisadores e colaboradores envolvidos no projeto, um controle no nível de usuário demandaria muito esforço para controlar os acessos a todos os conteúdos, já que exigiria a definição dos níveis de acesso individualmente para cada novo dado gerado.

A organização de pastas do Armazém de Dados também garante mais eficiência na gestão dos acessos, pois permite a atribuição do nível de acesso à pasta para um determinado grupo. Assim, todo o conteúdo armazenado herdará os mesmos níveis de acessos configurados para pasta ou para a estrutura a qual o dado pertence.

Este mesmo padrão, baseado em grupos de usuários e usuários relacionados, será utilizado por todas as ferramentas contratadas para as camadas de gestão de dados (Armazém de Dados, Repositório de Dados e Banco de Dados) do Programa de Monitoramento.

3.6.2 Existem restrições quanto ao compartilhamento de dados?

Ao compartilhar o conteúdo, é importante considerar as ações a seguir durante a fase de preparação dos dados.

- Formatação dos conjuntos de dados visando o acesso a longo prazo: O formato de arquivo no qual os seus dados são armazenados é um fator primordial na capacidade de reutilização futura. É necessário se planejar para a obsolescência de hardware e software. Veja formatos de arquivo e organização para detalhes sobre formatos de armazenamento de longo prazo. Na seção COLETA DE DADOS são apresentados os formatos recomendados que devem ser utilizados no PMBA.

- Documentação dos dados: Documentar a pesquisa e seus dados para que outros possam interpretar os dados. A documentação dos dados deve ser realizada desde o início do projeto de pesquisa e

continuar durante todo o projeto. Na seção DOCUMENTAÇÃO E METADADOS são apresentados maiores detalhes sobre a documentação de dados e metadados para o PMBA.

- Proteção e segurança: Verificar as implicações do compartilhamento de dados em termos de direitos autorais, propriedade intelectual e confidencialidade do assunto. Na seção ÉTICA E DIREITOS AUTORAIS são apresentados maiores detalhes sobre os direitos autorais e confidencialidade.

É terminantemente proibido enviar ou compartilhar dados confidenciais por pen drives, e-mail, Skype, Whatsapp e aplicações similares. O compartilhamento de dados deve ser sempre realizado por ferramentas especializadas, como o Dropbox, que protegem os dados com contas de usuários e senhas e possuem a funcionalidade de rastreabilidade habilitada.

Evite gerar dados criptografados. Eles podem ser efetivamente perdidos se forem criptografados com uma chave perdida (por exemplo, uma senha esquecida). Por esse motivo, as representações de dados criptografados são fortemente desencorajadas. Tenha em mente que as ferramentas de compartilhamento oficiais do projeto (Dropbox, Metacat, etc.) já utilizam métodos de encriptação e decifração dos dados de forma transparente, evitando que usuários não autenticados tenham acesso ao conteúdo das informações, mesmo que estas sejam interceptadas durante transferências.

Outras restrições podem ser devidas a acordos de confidencialidade e falta de consentimento dos órgãos financiadores do PMBA. Portanto, em caso de dúvidas sobre a publicação de qualquer material, a UGD recomenda uma consulta prévia ao NAIR, ESCPRO ou CTEC para análise.

3.6.3 Como dar o crédito adequado ao produtor dos dados?

Para dar crédito aos produtores dos dados é necessário utilizar os recursos de citação dos dados comumente oferecidos pelos Repositórios de Dados. A citação de dados é importante para:

- Conceder crédito adequado ao produtor de dados.
- Permitir acesso mais fácil aos dados para reutilização.
- Permitir que os leitores verifiquem seus resultados e análises.

Um conjunto de dados deve ser citado formalmente na lista de referências de um artigo, não apenas informalmente no texto. Muitos repositórios de dados, como o Metacat, e outros editores fornecem instruções explícitas para citar seus conteúdos. Abaixo veremos os elementos de citação que devem ser utilizados no cadastramento de um novo conjunto de dados ou data set no Repositório de Dados do PMBA (Metacat).

Elementos Centrais de Citação:

Existem 5 elementos principais que devem ser incluídos em uma citação do conjunto de dados.

- Criador (es): podem ser indivíduos ou organizações.
- Título: identificar textual para auxiliar nas buscas sobre os dados.

- Ano de publicação: quando o conjunto de dados foi lançado (pode ser diferente da data de acesso).
- Publicador: identifica o responsável pelo armazenamento e publicação.
- Identificador: um identificador público exclusivo (por exemplo, um ARK ou DOI).

Identificadores persistentes devem ser aplicados para que as pessoas possam encontrar seus dados de maneira confiável e eficiente. Eles também ajudam você a rastrear citações e a reutilizar o conteúdo.

Elementos Adicionais Comuns de uma Citação:

Embora os elementos principais sejam suficientes nos casos mais frequentes (citação para conjunto de dados estático), elementos adicionais podem ser necessários para citar um conjunto de dados dinâmico ou um subconjunto de um conjunto de dados maior.

- Versão: permite gerenciar o histórico de atualizações sobre o conjunto de dados.
- Data de acesso: quando os dados foram acessados .
- Subconjunto: intervalo de datas ou números de registro.
- Verificador de integridade: permite analisar se o conteúdo analisado é idêntico ao disponibilizado pelo autor (por exemplo, um Checksum).
- Localização na internet: necessário se o identificador não for "acionável" (passível de conversão para um endereço da web).

3.7 RESPONSABILIDADES E RECURSOS

Para se ter uma Gestão de Dados efetiva, é necessário mapear todas funções e atividades a serem desempenhadas e dividir as responsabilidades entre as equipes. É muito importante atribuir as funções de acordo com o conhecimento técnico de cada equipe. E em caso de deficiências na utilização das ferramentas de gestão de dados, devem ser realizadas capacitações e treinamentos para nivelamento.

Nesta seção serão abordadas questões relacionadas a divisão de responsabilidades entre os integrantes do PMBA. Também serão identificados todos os recursos necessários para a implementação do Plano de Gestão de Dados no projeto, incluindo treinamentos, recursos humanos, recursos de software e recursos de hardware.

3.7.1 Quem será responsável pelo gerenciamento de dados?

A responsabilidade por coordenar a implementação do Plano de Gestão de Dados, com todos os processos e recursos definidos neste documento, é da Unidade de Gestão de Dados (UGD) do PMBA. Porém, vários dos processos necessários para Gestão de Dados serão de responsabilidade de outras

equipes e terão sempre o suporte da UGD em caso de dúvidas ou problemas relacionados às ferramentas e metodologias recomendadas.

Abaixo são apresentadas as funções a serem desempenhadas pela Unidade de Gestão de Dados e as demais atividades, que serão de responsabilidade das equipes de cada Anexo/Projeto, contando sempre com o apoio da UGD.

Funções de Responsabilidade da UGD:

- Organização de Conteúdo: Definição e revisão, sempre que necessário, dos padrões de criação de pastas no Armazém de Dados, padrões de criação de metadados a serem disponibilizados no Repositório de Dados e modelagem de dados da camada de Banco de Dados.
- Organização de Usuários: Definição e revisão, sempre que necessário, das regras de criação de usuários e grupos de usuários nas ferramentas de Gestão de Dados.
- Monitoramento dos Recursos: Acompanhamento periódico de utilização dos recursos (humanos, hardware e software) contratados para o PMBA, identificando a necessidade de revisões, manutenções e novas aquisições.
- Treinamentos e Capacitações: Preparação de material de treinamento e realização de sessões de capacitação para o uso das ferramentas de gestão de dados contratadas para o projeto.
- Aquisição de Recursos de Gestão de Dados: Pesquisa e avaliação de opções do mercado para aquisição de novas ferramentas e recursos de apoio à Gestão de Dados adaptados aos requisitos do projeto.
- Instalação e Configuração de Recursos de Gestão de Dados: Atividades de ativação dos recursos adquiridos para o projeto relacionados a Gestão de Dados. Por exemplo, instalação e configuração dos servidores de dados e serviços digitais necessários para implementação das camadas de Gestão de Dados (Dropbox, Metacat, Morpho, PostgreSQL, etc.).
- Suporte às Ferramentas de Gestão de Dados: Acompanhamentos dos usuários e apoio técnico nas atividades de gestão de dados desempenhadas pelas equipes de Anexos/Projetos, como armazenamento, documentação e publicação de dados utilizando as ferramentas disponibilizadas pelas camadas de dados.
- Backups de Dados: Realização periódica das cópias de segurança das informações presentes nas camadas de dados do PMBA (Armazém de Dados, Repositório de Dados e Banco de Dados).

Funções de Responsabilidade das Equipes de Anexos/Projetos:

- Coleta de Dados: Aquisição dos dados em campo e digitalização dos resultados utilizando as planilhas padronizadas e formatadas pela CTEC e UGD.

- Armazenamento de Dados: Disponibilização dos dados e documentos gerados no Armazém de Dados (Dropbox) e no Repositório de Dados (Metacat), seguindo os padrões e organização de conteúdo estabelecidos pela UGD.
- Documentação e Metadados: Produção de metadados relacionados aos dados gerados pelo Anexo/Projeto utilizando a ferramenta Morpho e seguindo os padrões estabelecidos pela UGD.
- Validação de Conteúdo: Revisão e validação da qualidade de dados dos conteúdos gerados internamente pelos colaboradores do Anexo/Projeto.
- Publicação de Conteúdo: Publicação e divulgação dos dados e metadados no Repositório de Dados.
- Compartilhamento de Dados: Compartilhamento dos arquivos de trabalho armazenados no Armazém de Dados (Dropbox) com demais integrantes do Anexo/Projeto para permitir a colaboração. Compartilhamento dos dados e metadados via Repositório de Dados para os demais Anexos/Projetos.

3.7.2 Quais recursos serão necessários para implementar o plano de gestão de dados?

Os recursos necessários para implementação do Plano de Gestão de Dados foram divididos nas quatro categorias a seguir.

Treinamentos e Capacitações:

A UGD prevê a realização de seis turmas de treinamento para habilitar os pesquisadores e demais colaboradores dos Anexos/Projetos nas ferramentas utilizadas para interação com a camada de Repositório de Dados.

Cada turma deverá ser composta por no máximo cinco participantes e terá duração de dois dias, com o seguinte conteúdo.

- Dia 1 (Conteúdo Teórico): Apresentação das ferramentas Morpho e Metacat; apresentação de todas as etapas necessárias para publicação de conjuntos de dados no Repositório de Dados; apresentação dos padrões de criação de dados e metadados do PMBA.
- Dia 2 (Conteúdo Prático): Sessão prática para criação de conjuntos de dados e metadados relacionados aos resultados de cada Anexo/Projeto. A ideia é que sejam publicados no Repositório de Dados casos reais de conjuntos de dados criados pelos pesquisadores do Anexo/Projeto, seguindo os padrões definidos para o PMBA. Estes conjuntos de dados (data sets) poderão ser, posteriormente, utilizados como referência para os demais pesquisadores e colaboradores dos Anexos/Projetos.

No total, serão treinados até 30 pesquisadores e colaboradores do PMBA de todos os Anexos/Projetos. Após os treinamentos, os participantes estarão preparados para replicar o conhecimento adquirido internamente em seus respectivos Anexos/Projetos.

Recursos Humanos:

A Unidade de Gestão de Dados prevê a utilização de duas equipes distintas para implementação do Plano de Gestão de Dados.

- Equipe de Implantação: Responsável pelas atividades de instalação da infraestrutura prevista para as camadas de gestão de dados, definição dos padrões, organização de usuários e conteúdo. Atuará nos primeiros 2 anos de projeto.

- 1 coordenador: líder da equipe, responsável pela gestão dos demais recursos, divisão de atividades e geração dos status periódicos.
- 1 analista de sistemas sênior: especialista em modelagem de dados e integração de dados, para atuar na implementação da camada de Banco de Dados.
- 1 analista de sistemas pleno: especialista em metadados e conjuntos de dados de biodiversidade, para atuar na implementação da camada de Repositório de Dados.
- 1 analista de sistemas júnior: apoiar as atividades técnicas desempenhadas pelos demais integrantes da equipe e suportar a camada de Armazém de Dados que já está implementada.

- Equipe de Suporte: Responsável pelas atividades de monitoramento da utilização dos recursos implementados pelo projeto, suporte aos usuários na utilização das camadas de dados implementadas e realização de backups periódicos dos dados. Atuará nos últimos 3 anos do projeto e após a saída da equipe de implementação.

- 1 coordenador: líder da equipe, responsável pela gestão dos demais recursos, divisão de atividades e geração dos status periódicos.
- 1 analista de sistemas sênior: especialista em banco de dados, com experiência em repositórios de dados, para atuar no suporte a essas camadas de dados.
- 1 analista de sistemas pleno: apoiar as atividades técnicas desempenhadas pelos demais integrantes da equipe e suportar a camada de Armazém de Dados que já está implementada.

Recursos de Software:

A UGD, conforme já descrito em seções anteriores, prevê a utilização das seguintes ferramentas computacionais (softwares) para a implementação das camadas de dados.

- Armazém de Dados:

Dropbox Business: contratada para o primeiro ano do PMBA. Será necessário renovar o licenciamento para os 4 anos seguintes, mantendo a mesma quantidade de usuários e armazenamento ilimitado.

- Repositório de Dados:

- Morpho: ferramenta de metadados gratuita, não exige licenciamento.
- Metacat: ferramenta de gestão de repositório de dados gratuita, não exige licenciamento.

- Banco de Dados:

- PostgreSQL: ferramenta de gestão de banco de dados gratuita, não exige licenciamento.
- Pentaho Data Integration (PDI): ferramenta de integração de dados gratuita, não exige licenciamento.

O sistema operacional Linux será utilizado nos servidores de dados do projeto. O Linux também é gratuito e não exige licenciamento para uso.

Recursos de Hardware:

A UGD prevê a utilização dos seguintes componentes de Hardware para a implementação das camadas de dados.

- Servidor de Dados Local:

- Processador: 2 CPUs de 2GHz, com no mínimo 8 núcleos/ 16 segmentos cada.
- Memória: 64 GBytes (2 memórias de 32 GBytes RDIMM).
- Armazenamento:
 - 2 Discos de 1TB 10K RPM SATA 12Gbps, em RAID 1 para o sistema operacional.
 - 6 Discos de 4TB 7.2K RPM SATA 12Gbps, em RAID 5 para dados (repositório de dados e banco de dados).
 - 2 Storages de 120TB (12 discos de 10TB 7.2K RPM SATA 12Gbps, em RAID 5 para armazém de dados).
- Networking (Infraestrutura de Rede):
 - 2 interfaces de rede 1Gb.
 - Switch inteligente gerenciado de 24 portas de 1 GbE e 2 portas SFP de 1 GbE.
- Fontes: Duas fontes de alimentação redundantes de 1100W.
- Resfriamento: Opção de ventiladores redundante.
- Garantia: 3 anos de assistência técnica no local após diagnóstico remoto.

- Servidor de Dados na Nuvem:

- Instância: 2 vCPU(s), 64 GB de RAM.
- Armazenamento: 5 discos (4TB)
 - 5 discos (4TB) gerenciados do sistema operacional.
 - Storage adicional de 180 TB.

4 INTEGRAÇÃO UGD E UAR

Na seção anterior foram detalhadas as principais funcionalidades e requisitos das camadas de gestão de dados previstas para o PMBA. Nesta seção, serão abordadas as funcionalidades e requisitos do

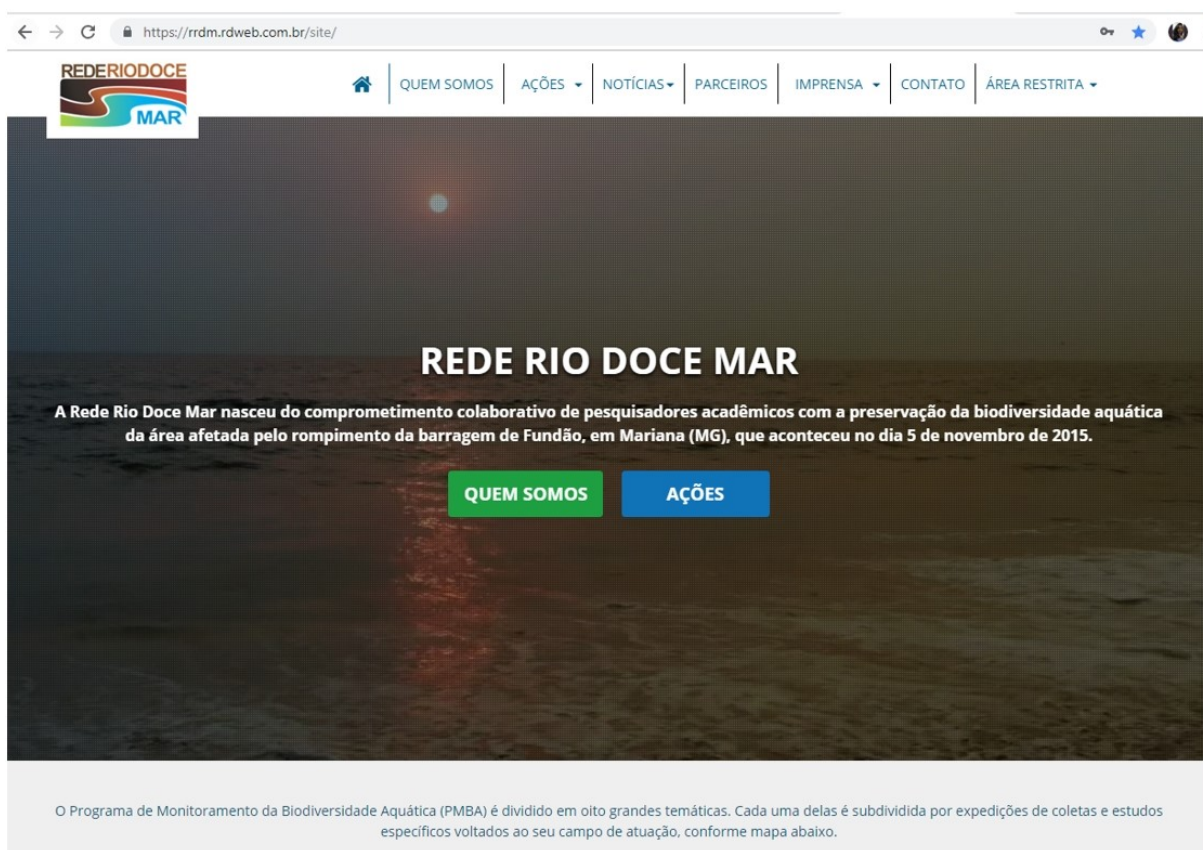
Sistema Web de Auxílio à Atuação Integrada em Rede (SWAR-RRDM), em desenvolvimento pela UAIR, e na sequência, serão apresentados os pontos de integração entre os produtos das duas Unidades do NAIR: UGD e UAR.

4.1 SISTEMA WEB DE AUXÍLIO À ATUAÇÃO EM REDE/RRDM (SW-A2R/RRDM)

O Sistema Web é o portal de comunicação da RRDM com a sociedade e com seus colaboradores.

O acesso ao portal pode ser realizado pelo link (<https://rrdm.rdweb.com.br/site/>). A Figura 14 apresenta a tela inicial do sistema.

Figura 14: Tela principal do Sistema Web da Rede Rio Doce Mar.



Fonte: autoria própria.

Nas próximas seções serão apresentadas as funcionalidades disponíveis para acesso pela sociedade e as funcionalidades restritas aos colaboradores da Rede Rio Doce Mar.

4.1.1 Informações Abertas à Sociedade

Na tela principal do sistema é possível acessar diversas informações relacionadas ao projeto, que estão disponíveis para o público geral, acessando um conjunto de itens de menus muito comuns em outros portais Web, como:

- Quem Somos: Apresentação da RRDM, sua origem e constituição.
- Ações: Divulgação das ações executadas, em andamento e futuras da RRDM.
- Notícias: Principais notícias e comunicados relacionados a atuação da RRDM.
- Parceiros: Informações relacionadas aos parceiros da RRDM e links para portais Web das instituições parceiras.
- Imprensa: Canal com o conteúdo da RRDM publicado na imprensa em geral.
- Contatos: Informações de localização e contato da RRDM.

4.1.2 Funcionalidades Restritas aos Colaboradores

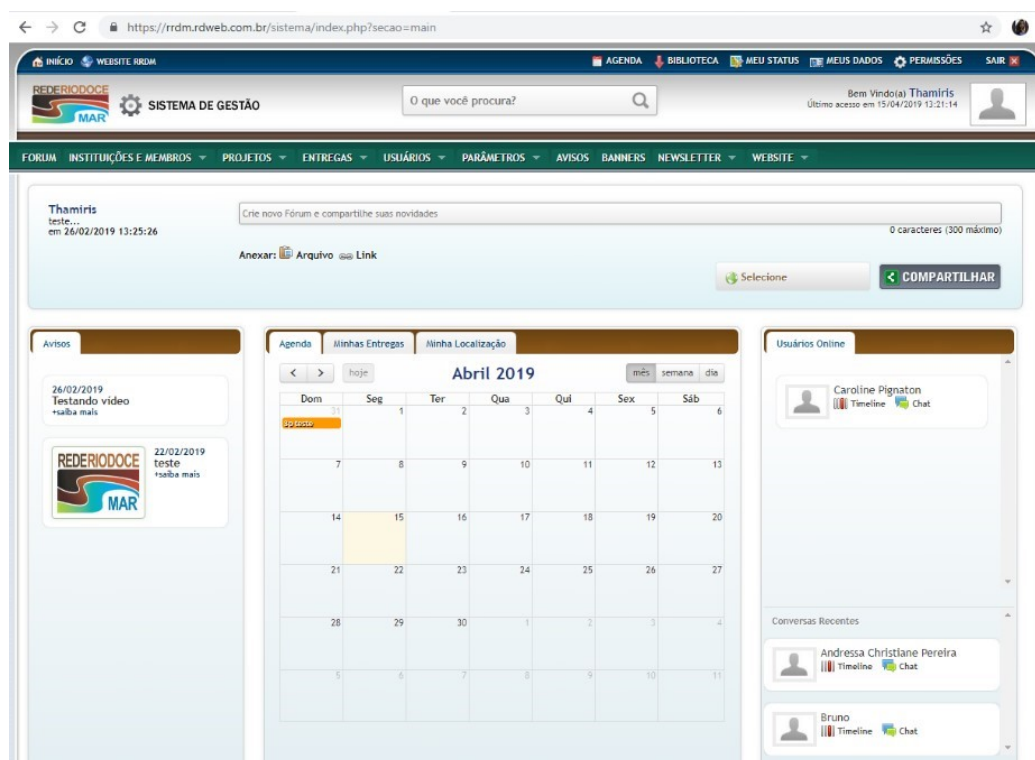
O sistema Web também possui uma área restrita aos colaboradores da RRDM. Para acessar esta área é necessário informar um e-mail e senha previamente cadastrado no sistema.

Após o acesso à área restrita, as seguintes funcionalidades ficam disponíveis na aplicação Web:

- Agenda: Cadastro e consulta de compromissos relacionados às atividades da RRDM.
- Fórum: Permite a participação em fóruns de discussões da RRDM.
- Biblioteca: Acervo dos documentos, imagens e vídeos compartilhados pelos usuários. É possível cadastrar os arquivos com links do dropbox.
- Meu status: Visualização da localização atual do colaborador. A localização também é compartilhada no Fórum do sistema. Esta função será mais bem aproveitada quando o sistema for adaptado para smartphones, pois facilitará o manuseio dos pesquisadores em campo.
- Banco de Contatos: Banco de dados dos colaboradores internos e externos da RRDM.
- Projetos e Entregas: Ferramenta para gestão das atividades em andamento durante o Programa. Acompanhamento e logística dos programas.
- Usuários: Área de cadastro e atualização dos usuários do sistema.
- Avisos e Banner: Ferramentas de comunicação interna na RRDM.

A Figura 15 apresenta a tela principal do sistema Web após acesso à área restrita da RRDM.

Figura 15: Tela principal da área restrita aos colaboradores da RRDM.



Fonte: autoria própria.

4.2 INTEGRAÇÃO DE AÇÕES E PRODUTOS

As equipes da UGD e da UAR realizaram um levantamento de todos os produtos e ações, relacionados a sistemas de informações previstos para o PMBA sob responsabilidade do NAIR. O resultado deste trabalho, detalhando cada uma das funcionalidades e atividades relacionadas às ferramentas de gestão de dados da UGD e ao sistema web da UAR, está disponível no anexo 39 (Análise de Integração de Sistemas do NAIR), disponibilizado na seção ANEXOS.

Com base neste levantamento, a equipe do NAIR identificou os pontos de integração entre os produtos e ações de cada unidade (UGD e UAR). Dentre os itens de integração identificados, podemos citar:

- Menu de Gestão de Dados: Criação de um grupo de itens (menu) no sistema web para centralizar todo o conteúdo relacionado a Gestão de Dados do PMBA, com os subitens a seguir.
- Documentação: Organização da documentação produzida e publicada pela Unidade de Gestão de Dados aos demais colaboradores da RRDM, como: manuais das aplicações; orientações relacionadas aos processos de gestão de dados (coleta de dados; geração metadados; curadoria de dados; armazenamento e backup; seleção e preservação; compartilhamento de dados; padronização de dados); modelos de planilhas de dados; modelos de dados e desenhos de arquitetura dos sistemas.
- Cronograma: Agenda de implementação das camadas de gestão de dados no PMBA e atividades

relacionadas, com reuniões e workshops com as equipes dos Anexos/Projetos.

- Suporte: Divulgação dos contatos e horários para suporte às aplicações e sistemas disponibilizados pela UGD. Integração com as funcionalidades de Fórum e Help Desk do sistema Web.
- Aplicações: Divulgação dos links para acesso às aplicações e sistemas disponibilizados pela UGD (Armazém de Dados; Repositório de Dados; Banco de Dados; e Análise de Dados).
- Treinamentos: Divulgação dos treinamentos e materiais de treinamento relacionados aos produtos implementados pela Unidade de Gestão de Dados.

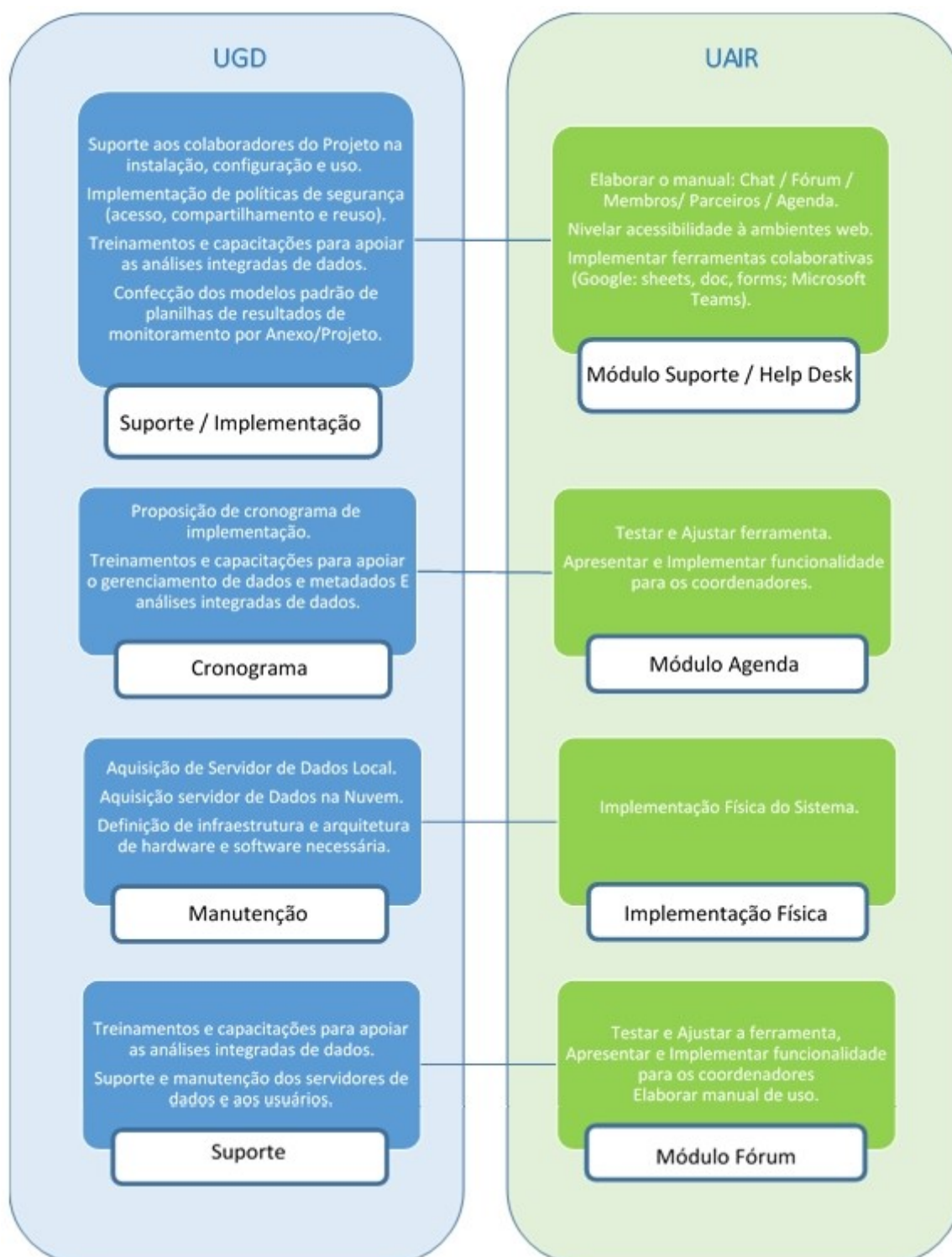
- Arquitetura de Gestão de Dados: Inclusão do Sistema Web no desenho de arquitetura de Gestão de Dados, considerando que o portal web passará a centralizar o acesso às demais aplicações previstas para implementação das camadas de gestão de dados do PMBA (Armazém de Dados; Repositório de Dados; Banco de Dados; e Análise de Dados).

- Compartilhamento de Infraestrutura: Unificação das infraestruturas de hardware, software e rede para os produtos computacionais das unidades (UGD e UAR). Assim, os sistemas informacionais do NAIR compartilharão a mesma infraestrutura.

- Compartilhamento de Equipe Técnica: Unificação das equipes técnicas que participarão da implantação e suporte dos produtos computacionais das unidades (UGD e UAR). Assim, os sistemas informacionais do NAIR compartilharão a mesma equipe técnica de tecnologia da informação.

A Figura 16 apresenta um fluxograma integrativo de ações e produtos entre as Unidades de Gestão de Dados (UGD) e de Atuação Integrada em Rede (UAR).

Figura 16: Fluxograma integrativo de ações e produtos entre as Unidades UGD e UAR.



Fonte: autoria própria.

Com base nos pontos de integração identificados, foi gerado um planejamento detalhado de ações a serem implementadas ao longo do PMBA, pela equipe do NAIR, para garantir a integração entre os produtos das unidades UGD e UAR. Este planejamento está disponível no anexo 40 (Plano de Integração de Sistemas do NAIR), disponibilizado na seção ANEXOS.

5 CUSTOS DE IMPLEMENTAÇÃO

Nesta seção serão apresentados os custos de infraestrutura (hardware e software) e de recursos humanos necessários para implementação das camadas de gestão de dados propostas para o PMBA. Nestes custos também estão incluídos os requisitos para implementação do Sistema Web de Auxílio à Atuação em Rede (SW-A2R/RRDM), em desenvolvimento pela UAR.

Como implementação, deve-se entender todas as atividades necessárias para garantir a operação das camadas de gestão de dados pelos seus usuários, compreendendo instalações, configurações, desenvolvimentos, manutenções, treinamentos e suporte técnico.

5.1 INFRAESTRUTURA

A seguir são apresentados os custos da infraestrutura de software e hardware necessários para o projeto.

A infraestrutura necessária para o Sistema Web da UAR (SW-A2R/RRDM) não exige grandes volumes de dados ou capacidade de processamento elevada, comparativamente com as camadas de gestão de dados. Sendo assim, podemos considerar que o Sistema Web (SW-A2R/RRDM) compartilhará a mesma infraestrutura de hardware e software dimensionada para as camadas de gestão de dados sem adição de novos custos.

5.1.1 Custos de Software

Como já comentado, o único custo de software relacionado às camadas de gestão de dados é do Dropbox Business para a camada de Armazém de Dados.

O Dropbox é licenciado anualmente. Para o primeiro ano do PMBA, o Dropbox foi contratado para 13 usuários com espaço de armazenamento ilimitado por US\$ 3.120,00.

A recomendação da Unidade de Gestão de Dados é que a contratação do Dropbox seja renovada para os próximos 4 anos, ao custo de US\$ 3.120,00 anuais ou US\$ 12.480,00.

O anexo 31 (Fatura Anual do Dropbox), disponibilizado na seção ANEXOS, apresenta a invoice (fatura) para contratação anual do Dropbox Business com espaço ilimitado para 13 usuários.

5.1.2 Custos de Hardware

A UGD prevê a utilização de dois servidores de dados para o projeto. Sendo um deles local e outro na nuvem, seguindo as configurações apresentadas na seção 3.7.2.

Abaixo são apresentados os custos de cada um dos servidores.

Servidor de Dados Local:

Em função do alto custo dos equipamentos, a UGD propõe a aquisição de um servidor com as mesmas configurações apresentadas anteriormente, porém com área de armazenamento suficiente apenas para implementação das camadas de Repositório de Dados e Banco de Dados. O anexo 32 (Cotação de Servidor Local - System), disponibilizado na seção ANEXOS, apresenta uma cotação, realizada no dia 14 de abril de 2019, no valor de R\$ 70.159,00. O servidor possui 2 TBytes em armazenamento do tipo RAID 1 (espelhamento), sendo 1 TByte útil para o sistema operacional e outros programas. Este tipo de armazenamento utiliza até 50% do espaço disponível para garantir a redundância e permitir a recuperação do sistema em caso de falhas. O servidor também possui 24 TBytes em armazenamento do tipo RAID 5, sendo 20 TBytes úteis para dados. Este tipo de armazenamento utiliza até 20% do espaço disponível para garantir a redundância e permitir a recuperação do sistema em caso de falhas.

As áreas de backups necessárias para Armazém de Dados, Repositório de Dados e Banco de Dados seriam inicialmente alocadas em pastas do Dropbox (já contratado). Após a finalização do projeto, caso a licença do Dropbox não seja renovada, será necessário adquirir uma infraestrutura de Storage (Armazenamento) equivalente a 180 TBytes úteis em armazenamento do tipo RAID 5. Este tipo de armazenamento utiliza até 20% do espaço disponível para garantir a redundância e permitir a recuperação do sistema em caso de falhas. Sendo assim, estima-se a necessidade de 240 TBytes de Storage. O anexo 33 (Cotação de Servidor Local - Storage), disponibilizado na seção ANEXOS, apresenta uma cotação, realizada no dia 14 de abril de 2019, de uma área de Storage de 120 TBytes, no valor de R\$ 122.461,00. Como seriam necessários 240 TBytes, o valor total estimado deve ser dobrado, sendo equivalente a R\$ 244.922,00.

Para compor a infraestrutura local, também foi realizada a cotação de um Switch inteligente com 24 portas no valor de R\$ 2.199,00. A cotação é do dia 14 de abril de 2019. Para maiores detalhes, favor consultar o anexo 34 (Cotação de Servidor Local - Networking), disponibilizado na seção ANEXOS.

Servidor de Dados na Nuvem:

Em função do alto custo da solução, a UGD propõe a aquisição de um servidor com as mesmas configurações apresentadas anteriormente, porém com área de armazenamento suficiente apenas para funcionamento do sistema operacional e instalação dos serviços de Repositório de Dados e Banco de Dados. No anexo 35 (Cotação de Servidor na Nuvem), aba "Inicial", disponibilizado na seção ANEXOS, é apresentada uma cotação, realizada no dia 14 de abril de 2019, no valor de R\$ 25.815,01 anuais. O servidor possui apenas 1 TByte em armazenamento para o sistema operacional e outros programas.

As soluções em nuvem são altamente escaláveis, o que permite que a infraestrutura possa ser ampliada automaticamente sem a necessidade de aquisição de equipamentos. Por isso, torna-se viável iniciar com um servidor com menor capacidade de armazenamento e ir ampliando o Storage (armazenamento) de acordo com o crescimento do projeto.

No anexo 35 (Cotação de Servidor na Nuvem), aba “Final”, disponibilizado na seção ANEXOS, também é apresentada uma cotação, realizada no dia 14 de abril de 2019, no valor de R\$ 370.235,24 anuais. O servidor possui a mesma área de armazenamento do servidor local apresentado anteriormente e só seria necessário após a finalização do projeto, caso a licença do Dropbox não seja renovada.

Vale lembrar que a solução em nuvem sugerida pela UGD não é vital para o projeto. O servidor em nuvem é apenas mais uma forma de garantir maior disponibilidade das soluções, por exemplo em caso de falhas nos servidores locais.

5.2 RECURSOS HUMANOS

A seguir são apresentados os custos de pessoal da Unidade de Gestão de Dados (UGD) necessários para o projeto.

Em função da grande sinergia entre as atividades relacionadas aos sistemas informacionais da UGD e da UAIR, aliada à necessidade de integração das ferramentas computacionais das unidades, foi dimensionada uma única equipe para implantação e suporte das soluções computacionais disponibilizadas pelo NAIR.

5.2.1 Equipe de Implantação

Como comentado, a equipe será responsável pelas atividades de ativação das camadas de gestão de dados e integração com o Sistema Web da UAIR (SWAR-RRDM), e atuará nos primeiros 2 anos de projeto.

Como já existe uma equipe definida e trabalhando no primeiro ano do projeto, a contratação da equipe de implantação proposta por este plano deverá ser efetivada apenas para o segundo ano do PMBA.

A Tabela 1 apresenta os custos estimados de cada profissional, com base em valores médios praticados pelo mercado. No total, são estimados R\$ 766.080,00 por ano para manutenção da equipe.

Tabela 1: Custos dos profissionais alocados na equipe de implantação

Profissional	Valor Hora (R\$)	Hora/mês	Custo Mensal (R\$)	Custo Anual (R\$)
Coordenador	\$140.00	168	\$23,520.00	\$282,240.00
Analista Sistemas Pleno	\$80.00	168	\$13,440.00	\$161,280.00
Analista Sistemas Sênior	\$120.00	168	\$20,160.00	\$241,920.00
Analista Sistemas Júnior	\$40.00	168	\$6,720.00	\$80,640.00
Total				\$766,080.00

Fonte: autoria própria.

Importante ressaltar que o Coordenador e o Analista de Sistemas Pleno deverão ser contratados em regime CLT pela RRDM. Os demais profissionais serão alocados, sob demanda, via contratação de horas com empresas terceiras especializadas em tecnologia da informação.

5.2.2 Equipe de Suporte

A equipe será responsável pelas atividades de suporte e manutenção das camadas de gestão de dados e Sistema Web da UAIR (SWAR-RRDM), e atuará nos últimos 3 anos de projeto.

A Tabela 2 apresenta os custos estimados de cada profissional, com base em valores médios praticados pelo mercado. No total, são estimados R\$ 558.720,00 anuais para manutenção da equipe.

Tabela 2: Custos dos profissionais alocados na equipe de suporte

Profissional	Valor Hora (R\$)	Hora/mês	Custo Mensal (R\$)	Custo Anual (R\$)
Coordenador	\$140.00	168	\$23,520.00	\$282,240.00
Analista Sistemas Pleno	\$80.00	168	\$13,440.00	\$161,280.00
Analista Sistemas Sênior	\$120.00	168	\$9,600.00	\$115,200.00
				\$558,200.00

Fonte: autoria própria.

Importante ressaltar que o Coordenador e o Analista de Sistemas Pleno deverão ser contratados em regime CLT pela RRDM. Os demais profissionais serão alocados, sob demanda, via contratação de horas com empresas terceiras especializadas em tecnologia da informação.

5.2.3 Treinamentos

Não estão sendo considerados custos adicionais para a realização dos treinamentos planejados, já que os próprios profissionais alocados na UGD serão responsáveis pelas atividades de confecção de material de treinamento e realização das sessões de capacitação com os pesquisadores e demais colaboradores do PMBA.

6 CRONOGRAMA DE IMPLEMENTAÇÃO

O Quadro 9 apresenta o cronograma de implementação das ações definidas neste Plano de Gestão de Dados para os cinco anos do Programa de Monitoramento PMBA. Para facilitar a sua leitura, cada ano foi dividido em trimestres.

Quadro 9: Cronograma de implementação do Plano de Gestão de Dados

Grupo	Atividade	Ano 1		Ano 2				Ano 3				Ano 4				Ano 5			
		T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Repositório de Dados de Testes	Instalação - Repositório de Dados de Testes	■																	
	Configuração - Repositório de Dados de Testes		■																
	Cargas - Repositório de Dados de Testes		■																
BD Data Lake de Testes	Modelagem - BD Data Lake de Testes	■																	
	Instalação - BD Data Lake de Testes	■																	
	Configuração - BD Data Lake de Testes		■																
	Cargas - BD Data Lake de Testes		■																
Ajustes e Aprovação do PGD	Ajustes e Aprovação do Plano de Gestão de Dados	■	■																
Aquisição Servidores de Dados	Aquisição de Servidor de Dados Local			■															
	Aquisição de Servidor de Dados na Nuvem			■															
Repositório de Dados Oficial	Instalação - Repositório de Dados Oficial				■														
	Configuração - Repositório de Dados Oficial					■													
	Cargas - Repositório de Dados Oficial						■	■	■	■	■	■	■	■	■	■	■	■	■
BD Data Lake Oficial	Modelagem - BD Data Lake Oficial			■	■														
	Instalação - BD Data Lake Oficial				■														
	Configuração - BD Data Lake Oficial					■													
	Cargas - BD Data Lake Oficial						■	■	■	■	■	■	■	■	■	■	■	■	■
BD Relacional Oficial	Modelagem - BD Relacional Oficial			■	■														
	Instalação - BD Relacional Oficial				■														
	Configuração - BD Relacional Oficial					■													

Grupo	Atividade	Ano 1		Ano 2				Ano 3				Ano 4				Ano 5			
		T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
	Cargas - BD Relacional Oficial																		
Integração UGD e UAIR	Publicação Conteúdo UGD no Sistema Web RRDM																		
	Integração Armazém de Dados e Sistema Web RRDM																		
	Integração Repositório de Dados e Sistema Web RRDM																		
	Integração Banco de Dados e Sistema Web RRDM																		
Treinamentos e Capacitações	Treinamentos e Capacitações																		
Suporte e Manutenção	Suporte e Manutenção dos Servidores de Dados																		
	Suporte aos Usuários das Camadas de Gestão de Dados																		
	Suporte aos Usuários do Sistema Web RRDM																		
	Apoio às Análises Integradas de Dados																		

Fonte: autoria própria.

É possível verificar que algumas atividades seguem sendo realizadas ao longo de todo o projeto: Suporte aos Usuários das Camada de Gestão de Dados; Suporte aos Usuários do Sistema Web RRDM; Suporte e Manutenção dos Servidores de Dados; Publicação de Conteúdo da UGD no Sistema Web RRDM; Cargas nas Camadas de Repositório de Dados e Banco de Dados; e Apoio às Análises Integradas de Dados. Estas atividades, em sua maioria de suporte e sustentação, são essenciais para garantir a utilização das camadas de gestão de dados pelos pesquisadores e demais colaboradores do PMBA. Um dos principais objetivos da UGD é suportar as Análises Integradas de Dados que serão realizadas pelas equipes dos Anexos/Projetos, e para isso é necessário estar em contato permanente com os pesquisadores, eliminando dúvidas, orientando, automatizando processos, padronizando e apoiando no uso das camadas de dados implementadas.

A seguir são apresentados detalhes adicionais sobre os planejamentos para os anos do cronograma, bem como equipes envolvidas e produtos esperados.

6.1 ANO 1 (ATUAL)

O principal produto da Unidade de Gestão de Dados (UGD) para o primeiro ano de monitoramento, ou ano atual, é o Plano de Gestão de Dados (PGD). Este produto, detalhado neste documento, está sendo entregue no final do segundo trimestre, e por isso apenas é apresentado o planejamento para os dois últimos trimestres do primeiro ano.

Além do Plano de Gestão de Dados, a UGD também já disponibilizou, nos dois primeiros trimestres de Ano 1, a camada de Armazém de Dados. Por isso não foram incluídas atividades referentes a implementação desta camada no cronograma.

As atividades previstas para o último semestre incluem a implementação de ambientes de testes para o Repositório de Dados e camada Data Lake do Banco de Dados. Estes ambientes serão configurados utilizando máquinas pessoais e não os servidores de dados previstos para o projeto. Por isso, estes ambientes não têm a pretensão de ter grande desempenho, nem de comportar toda a carga de dados do PMBA. A sua principal função será de testar os roteiros de instalação e configuração das ferramentas que serão utilizadas (Morpho, Metacat, PostgreSQL e PDI) e apoiar a definição dos padrões de organização de conteúdo, modelagem de dados e procedimentos de carga a serem seguidos pelas equipes dos Anexos/Projetos.

A implementação das camadas de dados oficiais, utilizando os servidores de dados estimados, foram planejados para o segundo ano do PMBA e após a aprovação do Plano de Gestão de Dados. Por isso, também é importante que todas as discussões e ajustes relacionados ao PGD sejam realizados ainda no primeiro ano, de forma a garantir uma versão aprovada para nortear as atividades do demais anos de projeto.

6.2 ANO 2

Enquanto o foco do primeiro ano está sendo o Planejamento, no segundo ano o foco da UGD será a Implementação das camadas de gestão de dados e o início da integração com o Sistema Web da RRDM. Este será o ano de ativação do Repositório de Dados e do Banco de Dados oficiais do PMBA.

De acordo com o planejamento, todos os equipamentos, ferramentas e serviços computacionais necessários para implementação das camadas de gestão de dados serão adquiridos ou contratados no primeiro trimestre. No segundo trimestre será iniciada a instalação e configuração das camadas. Atividades como modelagem de dados das camadas de Banco de Dados e definição dos padrões para organização de conteúdo no Repositório de Dados deverão ser finalizados até o final do terceiro trimestre. Todo o material de treinamento também deverá ser criado neste período.

No último trimestre do ano deverão ser iniciadas as capacitações das equipes dos Anexos/Projetos para que possam realizar as cargas de dados nos ambientes oficiais. As cargas dos conjuntos de dados e respectivos metadados deverão seguir por toda a extensão do projeto.

6.3 ANOS 3, 4 E 5

O foco dos três últimos anos é a manutenção dos ambientes oficiais de dados e apoio aos usuários no consumo das informações, principalmente para a geração das Análises Integradas de Dados. Também é esperado, para o terceiro ano, a finalização da integração das camadas de gestão de dados com o Sistema Web da RRDM.

A equipe da UGD seguirá apoiando as equipes de Anexos/Projetos na carga das informações no Repositório de Dados. A Unidade de Gestão de Dados também será responsável pela carga dos dados nas camadas do Banco de Dados, principal fonte para geração das análises integradas.

Outras atividades, como manutenção e monitoramento dos servidores de dados, gestão de acessos aos ambientes e realização de backups periódicos dos dados também serão de responsabilidade da equipe da UGD.

7 CONSIDERAÇÕES FINAIS

Como considerações finais, são apresentados alguns pontos de atenção que merecem monitoramento ao longo do projeto e algumas recomendações para o sucesso da Gestão de Dados no PMBA.

- Aprovação do PGD: O processo de validação e aprovação do Plano de Gestão de Dados pode ser demorado em função da grande quantidade de informações presentes na documentação. Mas é importante que a aprovação seja realizada ainda no final do primeiro ano de monitoramento para evitar

impactos no planejamento dos demais anos e na implementação do Repositório de dados e do Banco de Dados oficiais.

- Ajustes do PGD: O Plano de Gestão de Dados deve ser entendido como um guia para orientação das ações de gestão de dados do PMBA. O documento é passível de ajustes e atualizações ao longo do tempo. Porém é importante que todas as alterações sejam previamente alinhadas com todas as equipes envolvidas para evitar impactos não identificados ao longo do projeto.

- Servidores de Dados: As estimativas realizadas para os servidores de dados foram baseadas em alinhamentos prévios com as equipes dos Anexos/Projetos. É importante, ao longo do projeto, reavaliar as estimativas com base nos volumes reais de dados gerados. Também é importante iniciar a aquisição dos servidores partindo de uma configuração mínima, porém escalável, de forma a minimizar os custos sem impedir o crescimento da solução no futuro.

- Servidor em Nuvem: Vale lembrar que a solução em nuvem sugerida pela UGD não é vital para o projeto. O servidor em nuvem é apenas mais uma forma de garantir maior disponibilidade das camadas de dados, por exemplo em caso de falhas nos servidores locais. É necessário avaliar, ao longo do projeto, se este custo adicional se justifica em função da necessidade de disponibilidade dos dados.

- Segurança da Informação: Todos os dados brutos coletados pela RRDM no âmbito do Programa de Monitoramento são encaminhados, a cada 3 meses, em paralelo, para Fundação Renova, CIF (Comitê Interfederativo) e CT-Bio (Câmara Técnica de Conservação e Biodiversidade). Todo o material que a RRDM encaminha à Fundação Renova e ao CIF/CT-Bio, quando aprovado, se torna público. Antes dos dados se tornarem públicos, caso seja identificada a necessidade ou interesse em compartilhamento e/ou divulgação de informações para sociedade, instituições e/ou pesquisadores externos ao projeto, deve ser solicitada autorização prévia ao CIF/CT-Bio, por intermédio da Fundação Renova ou diretamente. O conteúdo das publicações não passa por censura da Fundação Renova.

- Documentação e Metadados: A documentação dos dados deve ser iniciada durante o planejamento da coleta e da geração dos dados, ou seja, ainda antes de se obter os dados. É importante ressaltar que, mesmo na ausência de uma ferramenta para apoiar a geração dos metadados (Morpho ou Metacat), eles podem ser gerados, por exemplo com a criação de arquivos texto a serem disponibilizados em conjunto com os arquivos de dados.

- Direitos Autorais: É importante que o PMBA defina, em comum acordo com Fundação Renova, CIF/CT-Bio e RRDM, qual padrão de licenciamento será utilizado para os dados gerados. Esta definição poderá ser realizada após a finalização do projeto, mas é importante que ocorra antes da publicação e divulgação dos dados a sociedade em geral.

8 ANEXOS

- ANEXO 01 – Estatuto da Fundação Renova.pdf: Documento oficial registrado em cartório com o Estatuto da Fundação Renova.
- ANEXO 02 – Apresentação RRDM.pdf: Apresentação sobre a RRDM abordando os tópicos: Origem; Modelo de Governança; Metodologia; Equipe Técnica.
- ANEXO 03 – Termo Monitoramento Cláusula 165 Geral.pdf: Termo de Referência do PMBA abordando os tópicos: Objetivos; Justificativas; Prazos.
- ANEXO 04 – Termo Monitoramento Cláusula 165 Anexo 2 Estudo Ambiente Dulcícolas.pdf: Anexo 02 do Termo de Referência do PMBA referente ao Estudo e Monitoramento do Ambiente Dulcícola.
- ANEXO 05 – Termo Monitoramento Cláusula 165 Anexo 3 Físico Químico.pdf: Anexo 03 do Termo de Referência do PMBA referente ao Estudo e Monitoramento no Rio Doce, Área Estuarina e Marinha.
- ANEXO 06 – Termo Monitoramento Cláusula 165 Anexo 4 Praias.pdf: Anexo 04 do Termo de Referência do PMBA referente ao Estudo e Monitoramento na Praia e Antepraia Adjacentes da Desembocadura do Rio Doce.
- ANEXO 07 – Termo Monitoramento Cláusula 165 Anexo 5 Manguezais.pdf: Anexo 05 do Termo de Referência do PMBA referente ao Estudo e Monitoramento dos Manguezais e Vegetação de Restinga sob Influência dos Sedimentos Provenientes do Rio Doce.
- ANEXO 08 – Termo Monitoramento Cláusula 165 Anexo 6 Megafauna.pdf: Anexo 06 do Termo de Referência do PMBA referente ao Estudo e Monitoramento da Megafauna.
- ANEXO 09 – Termo Monitoramento Cláusula 165 Anexo 7 Ictiofauna Marinha.pdf: Anexo 07 do Termo de Referência do PMBA referente ao Estudo e Monitoramento da Ictiofauna Marinha e Estuarina.
- ANEXO 10 – Termo Monitoramento Cláusula 165 Anexo 8 Abrolhos.pdf: Anexo 07 do Termo de Referência do PMBA referente ao Estudo e Monitoramento da Sedimentação no Parque Nacional Marinho dos Abrolhos e Regiões Relacionadas.
- ANEXO 11 – Plano Trabalho Geral.pdf: Plano de Trabalho Geral do PMBA abordando os tópicos: Objetivos; Justificativas; Equipe Executora; Produtos Esperados.
- ANEXO 12 – Plano Trabalho Anexo 1 Ecotoxicologia.pdf: Plano de Trabalho do Anexo/Projeto 01 do PMBA abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.
- ANEXO 13 – Plano Trabalho Anexo 3 Dulcícola.pdf: Plano de Trabalho do Anexo/Projeto 03 Dulcícola do PMBA abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.

- ANEXO 14 – Plano Trabalho Anexo 3 Marinho-Estuarino.pdf: Plano de Trabalho do Anexo/Projeto 03 Marinho-Estuarino do PMBA abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.
- ANEXO 15 – Plano Trabalho Anexo 4 Praias.pdf: Plano de Trabalho do Anexo/Projeto 04 do PMBA abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.
- ANEXO 16 – Plano Trabalho Anexo 5 Manguezal-Restinga.pdf: Plano de Trabalho do Anexo/Projeto 05 do PMBA abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.
- ANEXO 17 – Plano Trabalho Anexo 6 Megafauna.pdf: Plano de Trabalho do Anexo/Projeto 06 do PMBA abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.
- ANEXO 18 – Plano Trabalho Anexo 7 Ictiofauna.pdf: Plano de Trabalho do Anexo/Projeto 07 do PMBA abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.
- ANEXO 19 – Plano Trabalho Anexo 8 Abrolhos.pdf: Plano de Trabalho do Anexo/Projeto 08 do PMBA abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.
- ANEXO 20 – Equipes PMBA.xlsx: Documento com as equipes técnicas, equipes de laboratório e equipes de campo de cada Anexo/Projeto do PMBA.
- ANEXO 21 – Plano Trabalho UGD.pdf: Plano de Trabalho da Unidade de Gestão de Dados (UGD) abordando os tópicos: Objetivos; Metas e Justificativas; Equipe Técnica; Produtos; Metodologia.
- ANEXO 22 – Apresentação Solução Dados.pdf: Apresentação do desenho de arquitetura da solução de gestão de dados do PMBA.
- ANEXO 23 – Apresentação Armazém Dados.pdf: Apresentação da camada de Armazém de Dados (Dropbox) do PMBA.
- ANEXO 24 – Apresentação Repositório Dados.pdf: Apresentação da camada de Repositório de Dados (Morpho, Metacat e DataONE) do PMBA.
- ANEXO 25 – Apresentação Banco Dados.pdf: Apresentação da camada de Banco de Dados (PostgreSQL) do PMBA e ferramentas auxiliares para integração e visualização de dados (PDI e Tableau).
- ANEXO 26 – Guia Utilização Dropbox.pdf: Guia de Utilização do Dropbox, documentando os passos necessários para configurar e iniciar o uso da solução.
- ANEXO 27 – Protocolo Amostragem CTEC.pdf: Documento, de autoria do CTEC em conjunto com Anexos/Projetos, com os procedimentos de amostragem, acondicionamento e transporte das amostras, bem como de registro de dados in situ.

- ANEXO 28 – Calendário Campanhas Coleta.xlsx: Documento com o calendário de cada uma das 12 campanhas de monitoramento previstas para o primeiro ano do PMBA, com abertura por Anexo/Projeto.
- ANEXO 29 – Parâmetros do Programa de Monitoramento.xlsx: Documento com a lista de parâmetros a serem monitorados por cada Anexo/Projeto, além de informações complementares como pesquisador e laboratório responsável.
- ANEXO 30 – Planilha de Dados Padrão.xlsx: Documento com exemplo de planilhas pré-formatada para divulgação dos resultados de cada parâmetro monitorado. O exemplo é do Anexo/Projeto 1 (Ecotoxicologia). Na aba “Procedimentos” são detalhadas as instruções para preenchimento da planilha de resultados pelos pesquisadores.
- ANEXO 31 – Fatura Anual do Dropbox.pdf: Documento de invoice (fatura) para contratação anual do Dropbox Business com espaço ilimitado para 13 usuários.
- ANEXO 32 – Cotação de Servidor Local – System.pdf: Documento com a cotação do servidor de dados local, realizada no dia 14 de abril de 2019.
- ANEXO 33 – Cotação de Servidor Local – Storage.pdf: Documento com a cotação de storage (armazenamento) para o servidor de dados local, realizada no dia 14 de abril de 2019.
- ANEXO 34 – Cotação de Servidor Local – Networking.pdf: Documento com a cotação de switch (equipamento de rede) para o servidor de dados local, realizada no dia 14 de abril de 2019.
- ANEXO 35 – Cotação de Servidor na Nuvem.xls: Documento com a cotação do servidor de dados na nuvem, realizada no dia 14 de abril de 2019. O documento apresenta uma versão inicial e outra final para o servidor, considerando o aumento da área de storage (armazenamento).
- ANEXO 36 – Licenças Creative Commons.pdf: Documento com o detalhamento das condições e opções de licenciamento fornecidas pela Creative Commons.
- ANEXO 37 – Política Dados PPBio.pdf: Documento com a Política de Dados do Programa de Pesquisa em Biodiversidade (PPBio) publicada no Diário Oficial da União.
- ANEXO 38 – Política Dados PELD.pdf: Documento com a Política de Dados do Programa de Pesquisa Ecológica de Longa Duração (PELD) definida por Resolução Normativa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).
- ANEXO 39 – Análise Integração Sistemas NAIR.xlsx: Documento com o detalhamento de cada uma das funcionalidades e atividades relacionadas às ferramentas de gestão de dados da UGD e ao sistema web da UAIR, utilizado como base para identificação dos pontos de integração entre produtos e ações do NAIR.
- ANEXO 40 – Plano Integração Sistemas NAIR.xlsx: Documento com o planejamento detalhado de ações a serem implementadas ao longo do PMBA, pela equipe do NAIR, para garantir a integração entre os sistemas de informações das unidades UGD e UAIR.