##### Gráfico, Gráfico de radar Descrição gerada automaticamente

Modelo de mapeamento de componentes de solução tecnológica

Versão para avaliação e refinamento

Versão 0.1 – 14/06/2024

Fundação Renova

Histórico de Alterações

|  |  |  |  |
| --- | --- | --- | --- |
| Versão | Data | Descrição | Autor |
| 0.1 | 14/06/2024 | Criação do artefato. | FINOR |

Sumário

[Introdução 4](#_Toc169257012)

[Visão Geral 4](#_Toc169257013)

[Contexto do Negócio 4](#_Toc169257014)

[Desafios do Projeto 4](#_Toc169257015)

[Expectativas do Projeto 5](#_Toc169257016)

[Usuários da Solução 6](#_Toc169257017)

[Desafios e Capacidades da Equipe 6](#_Toc169257018)

[Especificação Técnica da Solução 9](#_Toc169257019)

[Análises e Desafios 9](#_Toc169257020)

[Fontes de Dados 10](#_Toc169257021)

[Formato de Dados 10](#_Toc169257022)

[Levantamento de Requisitos 11](#_Toc169257023)

[Levantamento de Requisitos do Modelo 12](#_Toc169257024)

[Plataformas de Análises 12](#_Toc169257025)

[Estruturas de Dados 12](#_Toc169257026)

[Confiabilidade dos Dados 12](#_Toc169257027)

# Introdução

# Visão Geral

## Contexto do Negócio

### Qual é o setor de atuação do cliente? Por exemplo, o cliente está envolvido no setor educacional, energético, saneamento ou em outro ramo específico?

O cliente é a Fundação Renova, a entidade responsável pela mobilização para a reparação dos danos causados pelo rompimento da barragem de Fundão, em Mariana (MG). [Fundação Renova - A Fundação](https://www.fundacaorenova.org/a-fundacao/).

A Fundação Renova tem como 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. [Fundação Renova - Estatuto](https://www.fundacaorenova.org/wp-content/uploads/2020/07/Estatuto-Registrado-2019.pdf).

### Quais são os principais produtos ou serviços oferecidos pela empresa em sua atividade principal?

O escopo da atuação da Fundação Renova, são 42 programas que se desdobram em diferentes projetos que estão sendo implementados nos 670 quilômetros de área impactada ao longo do rio Doce e afluentes. [Fundação Renova - A Fundação](https://www.fundacaorenova.org/a-fundacao/)

Os programas estão divididos em 3 eixos temáticos que buscam realizar reparações no contexto das pessoas e comunidades, terra e água (gestão de solo, recursos hídricos, biodiversidade), e de reconstrução e infraestrutura. [Fundação Renova - Conheça os programas](https://www.fundacaorenova.org/conheca-os-programas/)

A Fundação Renova é responsável pela gestão desses programas e pela publicização dos resultados obtido com esses programas, submetendo estes resultados ao Comitê Interfederativo (CIF) - órgão presidido pelo IBAMA para validar as medidas de reparação. [Ibama - CIF](https://www.ibama.gov.br/cif)

## Desafios do Projeto

### Quais são os desafios relacionados a dados enfrentados atualmente pela empresa que motivaram a contratação de um projeto?

1. Desafios pacificados:

* Diferentes fornecedores de dados.
* Dificuldade com padronização dos dados.
* Dificuldade de reunir e dar acesso livre aos dados produzidos para a sociedade.
* [Evidência: Weekly PMQQS - 11/06/204 - 30:40](https://nasnuvensrnp-my.sharepoint.com/personal/rosana_caetano_consultores_rnp_br/_layouts/15/stream.aspx?id=%2Fpersonal%2Frosana%5Fcaetano%5Fconsultores%5Frnp%5Fbr%2FDocuments%2FRecordings%2F%5BRenova%5D%20%2D%20Weekly%20PMQQS%2D20240611%5F160304%2DGrava%C3%A7%C3%A3o%20de%20Reuni%C3%A3o%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Ebe967500%2D32f5%2D45f8%2D8c8c%2D23bdb2db6464&ga=1)

1. Desafios em discussão:

* Ausência de um repositório/banco de dados comum para todos os dados.

### O cliente enfrenta alguma dificuldade específica em relação à gestão ou uso de dados atualmente? Em caso afirmativo, como essa dificuldade pode representar um entrave para o projeto

* Grande esforço manual para processamento dos arquivos.
* Ausencia de uma estrutura de armazenamento comum para dados de um mesmo domínio (e.g. dados bióticos).
* Ausência de padronização das fontes de dados. Há formatos distintos (e.g. .csv, .xlsx, .txt).
* Ausência de relações definidas entre dados de um mesmo domínio.
* Dificuldade de reunir esses dados e disponibilizar os dados obtidos pelos programas para uso dos órgãos governamentais e da sociedade.
* [Evidência: Weekly PMQQS - 11/06/204 - 30:40](https://nasnuvensrnp-my.sharepoint.com/personal/rosana_caetano_consultores_rnp_br/_layouts/15/stream.aspx?id=%2Fpersonal%2Frosana%5Fcaetano%5Fconsultores%5Frnp%5Fbr%2FDocuments%2FRecordings%2F%5BRenova%5D%20%2D%20Weekly%20PMQQS%2D20240611%5F160304%2DGrava%C3%A7%C3%A3o%20de%20Reuni%C3%A3o%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Ebe967500%2D32f5%2D45f8%2D8c8c%2D23bdb2db6464&ga=1)

### Houve alguma iniciativa interna anterior que abordou os desafios que serão enfrentados por este projeto? Em caso afirmativo, o cliente poderia compartilhar os aprendizados adquiridos com essa iniciativa?

*[Finor - É necessário coletar evidências junto ao time da Clara e dos analistas de negócio.]*

## Expectativas do Projeto

### Quais são, nas palavras do cliente, os principais objetivos deste projeto de dados?

* São prioridades, nas palavras da Andreza Malini Gonçalves (Renova), para os dados Bióticos/Biodiversidade:

1. Conseguir tratar os dados, ou seja, transformá-los para uma forma padronizada.
2. Dar visibilidade para esses dados utilizando o IPT/SiBBr > [Weekly PMQQS 11/06/204 - 24:08](https://nasnuvensrnp-my.sharepoint.com/personal/rosana_caetano_consultores_rnp_br/_layouts/15/stream.aspx?id=%2Fpersonal%2Frosana%5Fcaetano%5Fconsultores%5Frnp%5Fbr%2FDocuments%2FRecordings%2F%5BRenova%5D%20%2D%20Weekly%20PMQQS%2D20240611%5F160304%2DGrava%C3%A7%C3%A3o%20de%20Reuni%C3%A3o%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Ebe967500%2D32f5%2D45f8%2D8c8c%2D23bdb2db6464&ga=1)

* A solução SiBBr/IPT foi contratada como uma alternativa para dar visibilidade a esses dados e disponibilizá-los para utilização da comunidade.

### O cliente definiu alguma métrica específica para avaliar o sucesso do projeto?

* Não há evidências da métrica de sucesso que será utilizada para determinar o sucesso do projeto.

# Usuários da Solução

## Desafios e Capacidades da Equipe

A equipe interna possui experiência anterior no desenvolvimento de projetos de dados, incluindo o uso de metodologias ágeis ou outras abordagens de desenvolvimento de software?

* Equipe interna não é formada por profissionais de tecnologia.
* Equipe é multidisciplinar formada de biólogos e engenheiros ligados à área ambiental.
* Alguns dos membros da equipe interna já contatados:
  + Andreza Malini Gonçalves - Engenheira Sanitária
  + Leandro Fernandes Antonio Santos - Cientista Biológico/Biólogo
  + Bianca Menezes Lima Cardoso - Cientista Biológica/Bióloga
  + Matheus Guimarães Cardoso Nogueira - Cientista Biológico/Biólogo

### Caso já tenha executado outros projetos de dados, o cliente poderia compartilhar os principais desafios enfrentados pela equipe interna ao executar com projetos de dados anteriores?

* O cliente possui iniciativas internas de dados.
* As dificuldades indicadas pelo cliente estão fundamentadas na falta de infraestrutura de gerenciamento de dados própria para os desafios enfrentados pela equipe.

*[Finor - Falta produzir evidências para basear essa resposta no contexto de V1]*

### A equipe interna do cliente possui profissionais especializados em dados, como Engenheiros de Dados, Cientistas de Dados, Analistas de Dados, Engenheiros de Machine Learning, Administradores de Banco de Dados (DBAs), entre outros? Em caso afirmativo, poderia fornecer informações sobre o tamanho atual da equipe?

* Não há indícios da existência de um time de Engenheiro de Dados e Cientistas de Dados próximos do time negocial (Biólogos da Fundação Renova).
* Os fluxos de dados operados hoje pelo cliente são realizados por técnicos que não possuem formação específica em gerenciamento de dados. Uma evidência é o gerenciamento das bases de dados do programa de monitoramento da agua descrito em [reunião](https://nasnuvensrnp-my.sharepoint.com/personal/rosana_caetano_consultores_rnp_br/_layouts/15/stream.aspx?id=%2Fpersonal%2Frosana%5Fcaetano%5Fconsultores%5Frnp%5Fbr%2FDocuments%2FRecordings%2F%5BRenova%5D%20%2D%20Weekly%20PMQQS%2D20240611%5F160304%2DGrava%C3%A7%C3%A3o%20de%20Reuni%C3%A3o%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Ebe967500%2D32f5%2D45f8%2D8c8c%2D23bdb2db6464&ga=1) que é operado por uma bióloga (Bianca Menezes Lima Cardoso).

### A equipe está familiarizada com ferramentas e tecnologias emergentes no campo da análise de dados, como aprendizado de máquina, big data, ou computação em nuvem?

*[Finor - É necessário coletar evidências junto ao time negocial para embasa a resposta a esse questionamento.]*

Usuários e Mantenedores da Solução

### **Quem serão os usuários da solução proposta? Esses usuários são internos ou externos à organização do cliente?**

* Os usuários da solução são internos e externos à organização:
  + **Administrador da Instituição:** É um usuário interno. Deverá ser capaz de realizar gestão e controle de acesso dos usuários, gerenciar os recursos de dados criados pelos usuários e também ser capaz de gerar relatórios gerenciais e técnicos.
  + **Fornecedores de dados:** São usuários externos, como laboratórios e pesquisadores, que realizam pesquisas de campo e posteriormente inserem os dados na solução.
  + **Visitante:** É um usuário externo que utiliza a solução da RNP para atestar a qualidade dos dados inseridos na plataforma. Trata-se de um auditor externo.
  + **Administrador RNP:** É um usuário externo responsável por dar suporte à solução construída para a empresa. Esse usuário é necessário, pois não há uma equipe de tecnologia capaz de realizar atualizações na plataforma construída pelo cliente. O administrador RNP deverá ser capaz de realizar a gestão do IAM (Identity and Access Management) da plataforma.

[Mapa de Atores Negociais](https://miro.com/app/board/uXjVKF2pf00=/)

### Seria viável que o cliente fornecesse uma descrição detalhada da persona que utilizará essa ferramenta, a fim de garantir uma abordagem mais alinhada às suas necessidades específicas?

*[Finor - É necessário coletar evidências junto ao time negocial para embasa a resposta a esse questionamento.]*

### Qual é o nível técnico ou familiaridade dos usuários finais em relação às áreas de engenharia, ciência e análise de dados? É possível fornecer informações sobre o grau de experiência que esses usuários têm nessas disciplinas específicas?

* Não há indícios da existência de um time de Engenheiro de Dados e Cientistas de Dados próximos do time negocial (Biólogos da Fundação Renova).
* Os fluxos de dados operados hoje pelo cliente são realizados por técnicos que não possuem formação específica em gerenciamento de dados. Uma evidência é o gerenciamento das bases de dados do programa de monitoramento da agua descrito em [reunião](https://nasnuvensrnp-my.sharepoint.com/personal/rosana_caetano_consultores_rnp_br/_layouts/15/stream.aspx?id=%2Fpersonal%2Frosana%5Fcaetano%5Fconsultores%5Frnp%5Fbr%2FDocuments%2FRecordings%2F%5BRenova%5D%20%2D%20Weekly%20PMQQS%2D20240611%5F160304%2DGrava%C3%A7%C3%A3o%20de%20Reuni%C3%A3o%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Ebe967500%2D32f5%2D45f8%2D8c8c%2D23bdb2db6464&ga=1) que é operado por uma bióloga (Bianca Menezes Lima Cardoso).

### **O cliente possui um setor dedicado à governança e segurança de dados? Esse setor está disponível para participar do processo de desenho do escopo do projeto?**

*[Finor - É necessário coletar evidências junto ao time negocial para embasa a resposta a esse questionamento.]*

# Especificação Técnica da Solução

## Análises e Desafios

### Há a necessidade de colaboração simultânea entre a equipe interna e a RNP? Caso positivo, existem diretrizes específicas estabelecidas para orientar essa colaboração?

*[Comentário Finor - Ainda cabem esclarecimentos sobre essa questão antes de fornecer uma resposta definitiva a esse questionamento].*

### **Existem políticas de segurança de dados que exigem que o desenvolvimento, manipulação e análise dos dados ocorram na infraestrutura interna do cliente?**

* Não há menção de políticas de segurança que atuem de forma impeditiva na manipulação de dados fora da infraestrutura do cliente.
* O projeto será desenvolvido em um ambiente de nuvem construído pela RNP.
* O ambiente de nuvem utilizado será multicloud com o webapp sendo hospedado na Amazon Web Services (AWS) e operações de manipulação e qualidade de dados hospedado no Google Cloud Plataform (GCP).

[Fonte - Solicitação de Proposta Comercial (011/2024) - Fundação Renova - Data](https://nasnuvensrnp.sharepoint.com/:b:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/Artefatos%20da%20Finor/V1/Assessment/assets/Renova_SGBD_Solicita%C3%A7%C3%A3o_de_Proposta_2024_011_SQUAD-DataQuality_V1.0%20-%20Assinada.pdf?csf=1&web=1&e=ovSIjX)

### O projeto será desenvolvido na infraestrutura computacional da empresa? Existe uma equipe interna capacitada para fornecer a infraestrutura necessária para o desenvolvimento do projeto?

* O projeto não será desenvolvido na infraestrutura do cliente.
* O projeto será desenvolvido em um ambiente de nuvem construído pela RNP.
* O ambiente de nuvem utilizado será multicloud.
* O webapp será hospedado na Amazon Web Services (AWS).
* As operações de manipulação e verificação de qualidade de dados serão hospedadas no Google Cloud Plataform (GCP).
* Não há indicios de que necessário a disponibilidade da equipe interna para gestão da infraestrutura necessária para desenvolver a solução.

[Fonte - Solicitação de Proposta Comercial (011/2024) - Fundação Renova - Data Quality](https://nasnuvensrnp.sharepoint.com/:b:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/Artefatos%20da%20Finor/V1/Assessment/assets/Renova_SGBD_Solicita%C3%A7%C3%A3o_de_Proposta_2024_011_SQUAD-DataQuality_V1.0%20-%20Assinada.pdf?csf=1&web=1&e=ovSIjX)

### Existe a necessidade de manipulação de grandes volumes de dados (big data) durante a execução do projeto?

* O projeto apresenta evidências de que não será necessária a manipulação de conjuntos de dados que se encaixam na definição de Big Data.
* Os arquivos de interesse podem ter entre 50KB a 3MB de tamanho e algo entre 200 a 20.000 registros.
* Estima-se que o volume de arquivos a ser ingerido seja por volta de 200 a 300 por semestre.

[RNP\_RENOVA\_Caso de Teste POC\_v1](https://nasnuvensrnp.sharepoint.com/:b:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/Artefatos%20da%20Finor/V1/Assessment/assets/Renova_SGBD_Solicita%C3%A7%C3%A3o_de_Proposta_2024_011_SQUAD-DataQuality_V1.0%20-%20Assinada.pdf?csf=1&web=1&e=ovSIjX)

### O projeto será desenvolvido na infraestrutura computacional da empresa? Existe uma equipe interna capacitada para fornecer a infraestrutura necessária para o desenvolvimento do projeto?

* O projeto não será desenvolvido na infraestrutura do cliente.
* O projeto será desenvolvido em um ambiente de nuvem construído pela RNP.
* O ambiente de nuvem utilizado será multicloud.
* O webapp será hospedado na Amazon Web Services (AWS).
* As operações de manipulação e verificação de qualidade de dados serão hospedadas no Google Cloud Platform (GCP).
* Não há indícios de que necessário a disponibilidade da equipe interna para gestão da infraestrutura necessária para desenvolver a solução.

[Fonte - Solicitação de Proposta Comercial (011/2024) - Fundação Renova - Data Quality](https://nasnuvensrnp.sharepoint.com/:b:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/Artefatos%20da%20Finor/V1/Assessment/assets/Renova_SGBD_Solicita%C3%A7%C3%A3o_de_Proposta_2024_011_SQUAD-DataQuality_V1.0%20-%20Assinada.pdf?csf=1&web=1&e=ovSIjX)

## Fontes de Dados

### Relacionado ao desafio a ser enfrentado, existe conhecimento sobre a necessidade de enriquecimento dos dados por meio de fontes externas à infraestrutura interna?

* Não é necessário o enriquecimento e/ou a contratação de serviços externos de dados para complementar as bases de dados do cliente.
* O Cliente se utiliza de laboratórios de pesquisa, universidades e outros agentes externos, como fornecedores de dados.
* Essa característica imputa desafios semelhantes ao uso de fontes externas pois traz desafios relacionados a padronização e volatilidade da qualidade dos dados
* É necessário cuidado adicional durante o processo de ingestão dos dados.

## Formato de Dados

### Existe um mapeamento prévio em relação aos formatos de dados necessários para a ingestão e integração?

* [x] CSV
  + - Não foram localizados arquivos CSV nas pastas de exemplos de arquivos de entrada.
* [x] Planilhas não estruturadas
  + - Planilha não estruturada de dados de localização geográfica de pontos de coleta. [Evidência: Pontos amostrais.xlsx](https://nasnuvensrnp.sharepoint.com/:x:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/csv%20exemplos/Bruto/PG28/164a_2019_Econservation/Pontos%20amostrais.xlsx?d=w6a5c9eb50c5145c391ed14e1851519bd&csf=1&web=1&e=jhIIcA)
    - Planilha não estruturada de dados bióticos (correspondente à planilha PLANIL~1.XLS do exemplo seguinte). [Evidência: MATRIZ~1.XLS](https://nasnuvensrnp.sharepoint.com/:x:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/csv%20exemplos/Bruto/PG28/165_2020_Cepemar/2020-05-20_Dados%20Brutos_Monitoramento_BiotaAqua%CC%81tica_Anexo2_Cl165/MATRIZ~1.XLS?d=wf58c013136904bab8672f68b951a431d&csf=1&web=1&e=u275QZ)
* [x] Planilhas em formato estruturado
  + - Planilha estruturada de dados bióticos (correspondente à planilha MATRIZ~1.XLS do exemplo anterior). [Evidência: PLANIL~1.XLS](https://nasnuvensrnp.sharepoint.com/:x:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/csv%20exemplos/Bruto/PG28/165_2020_Cepemar/2020-05-20_Dados%20Brutos_Monitoramento_BiotaAqua%CC%81tica_Anexo2_Cl165/PLANIL~1.XLS?d=w265a2b8dc2b741a9b164a11db0b5430b&csf=1&web=1&e=CtL5iK)
* [ ] Parquet
* [ ] Bancos de dados relacionais
* [ ] JSON
* [ ] XML
* [ ] PDF
* [ ] Documentos de texto não estruturados
* [ ] Áudio
* [ ] Imagem
* [ ] Vídeo

As evidências anexadas acima são provenientes das seguintes pastas com exemplos de arquivos de entrada: [PG28](https://nasnuvensrnp.sharepoint.com/:f:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/csv%20exemplos/Bruto/PG28?csf=1&web=1&e=IFbT7j) e [PG30](https://nasnuvensrnp.sharepoint.com/:f:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/csv%20exemplos/Bruto/PG30?csf=1&web=1&e=MWErde)

## Levantamento de Requisitos

### O cliente espera algum dos seguintes resultados com o projeto?

[x] Disponibilidade de dados para consumo.

[x] Reduzir dificuldade no acesso dos dados.

[x] Melhorar o entendimento dos dados.

[x] Melhorar a visibilidade dos dados.

[ ] Análises com domínio específico dos dados.

[ ] Análises estatísticas dos dados.

[ ] Exploração dos dados além das estratégias descritivas, de visualização e de estatísticas.

[ ] Entender ou explicar os impactos de fatores sobre alguma métrica.

[ ] Automações de decisões e previsões de dados.

[ ] Automações de alertas e ações com base nos dados.

[ ] Automações de recomendações e classificações com base nos dados.

[ ] Atingir conformidade com regulamentações de privacidade de dados, como LGPD, GDPR ou HIPAA

[ ] Implementar medidas específicas para proteger informações pessoais identificáveis (PII)

* Cliente não está interessado que o projeto realize análises estatísticas e descritivas com os dados.
* Cliente não está interessado em criar modelos de automação de decisão.
* Não está claro se o cliente tem interesse em ser capaz de acessar as bases de dados que foram estruturadas durante o processo de ingestão e armazenamento [Evidência: Weekly PMQQS](https://nasnuvensrnp-my.sharepoint.com/personal/rosana_caetano_consultores_rnp_br/_layouts/15/stream.aspx?id=%2Fpersonal%2Frosana%5Fcaetano%5Fconsultores%5Frnp%5Fbr%2FDocuments%2FRecordings%2F%5BRenova%5D%20%2D%20Weekly%20PMQQS%2D20240611%5F160304%2DGrava%C3%A7%C3%A3o%20de%20Reuni%C3%A3o%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Ebe967500%2D32f5%2D45f8%2D8c8c%2D23bdb2db6464&ga=1).
* Não está claro que o cliente quer ter acesso direto aos serviços de nuvem que armazenam os dados [Evidência: Weekly PMQQS](https://nasnuvensrnp-my.sharepoint.com/personal/rosana_caetano_consultores_rnp_br/_layouts/15/stream.aspx?id=%2Fpersonal%2Frosana%5Fcaetano%5Fconsultores%5Frnp%5Fbr%2FDocuments%2FRecordings%2F%5BRenova%5D%20%2D%20Weekly%20PMQQS%2D20240611%5F160304%2DGrava%C3%A7%C3%A3o%20de%20Reuni%C3%A3o%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Ebe967500%2D32f5%2D45f8%2D8c8c%2D23bdb2db6464&ga=1).

## Levantamento de Requisitos do Modelo

As questões desse subconjunto não se aplicam pois o projeto não tem um foco de ciência de dados.

## Plataformas de Análises

As questões desse subconjunto não se aplicam pois o projeto não tem um foco de ciência de dados.

# Estruturas de Dados

## Confiabilidade dos Dados

### Há análises e monitoramento da qualidade dos dados, em especial os potencialmente relacionados à solução de dados?

* Atualmente, há análise, monitoramento e correção da qualidade dos dados manualmente, por parte da equipe técnica de biólogos. [Evidência: Weekly PMQQS](https://nasnuvensrnp-my.sharepoint.com/personal/rosana_caetano_consultores_rnp_br/_layouts/15/stream.aspx?id=%2Fpersonal%2Frosana%5Fcaetano%5Fconsultores%5Frnp%5Fbr%2FDocuments%2FRecordings%2F%5BRenova%5D%20%2D%20Weekly%20PMQQS%2D20240611%5F160304%2DGrava%C3%A7%C3%A3o%20de%20Reuni%C3%A3o%2Emp4&referrer=StreamWebApp%2EWeb&referrerScenario=AddressBarCopied%2Eview%2Ebe967500%2D32f5%2D45f8%2D8c8c%2D23bdb2db6464&ga=1).

### Há conhecimento de inconsistências nos dados, especialmente os relacionados ao desafio da solução de dados?

* Há conhecimento de inconsistências nos dados, o que constitui um dos desafios do projeto.
* Um exemplo de inconsistência está nos arquivos brutos de coleta de dados bióticos que é fornecido com os cabeçalhos incompatíveis com as regras DarwinCore. [Evidência: MATRIZ~1.XLS](https://nasnuvensrnp.sharepoint.com/:x:/r/sites/RenovaCooperaoFundaoRenova/Biblioteca%20de%20Execuo/2.%20Execu%C3%A7%C3%A3o/2.10.%20Documentos%20gerais/DocumentosTecnicos_CompartilhadosFornecedores/csv%20exemplos/Bruto/PG28/165_2020_Cepemar/2020-05-20_Dados%20Brutos_Monitoramento_BiotaAqua%CC%81tica_Anexo2_Cl165/MATRIZ~1.XLS?d=wf58c013136904bab8672f68b951a431d&csf=1&web=1&e=dMo19w)
* A solução precisa ser capaz de identificar inconsistências nos arquivos de dados brutos e oferecer uma maneira de corrigir esses arquivos.

### Há documentação de processamentos que tratam inconsistências nos dados, especialmente os relacionados ao desafio da solução de dados?

* Há documentação de processamentos que tratam inconsistências nos dados, conhecidos como regras Darwin Core, publicamente acessíveis e mantidas por uma comunidade ativa de contribuintes do mundo todo.
* Exemplo de documentação pode ser encontrado no repositório TDWG no GitHub. [Evidência: Core Tests and Assertions (TG2)](https://github.com/tdwg/bdq/projects/2)
* Exemplo de documentação pode ser acessado no repositório DarwinCoreBrasil no GitHub. [Evidência: Testes de qualidade de dados](https://github.com/sibbr/DarwinCoreBrasil/blob/regras-validacao/regras-validacao/regras.md)