

BIG DATA E GOVERNANÇA AMBIENTAL: PROJETO INTERNACIONAL DE MONITORAMENTO DA QUALIDADE DA ÁGUA EM ESCALA NACIONAL

Ana Carolina Santoro¹ (ana.santoro@waterservicestech.com), Danilo de Castro Silva², Aschalew Debebe³

Resumo

Programas nacionais de monitoramento ambiental enfrentam desafios significativos na consolidação e gestão de grandes volumes de dados heterogêneos. Em parceria com o órgão regulador ambiental local, este projeto contemplou o desenvolvimento de uma infraestrutura robusta de dados para embasar a elaboração do Plano Nacional de Monitoramento de Qualidade Hídrica de uma região do Oriente Médio, a partir do diagnóstico da qualidade hídrica e do mapeamento de áreas prioritárias para investigação de contaminação hídrica. A metodologia combinou a estruturação e validação de um extenso banco de dados histórico, usando métodos e processos da engenharia de dados, com a execução de uma nova campanha nacional de amostragem, destinada a preencher lacunas espaciais e temporais identificadas. A arquitetura de dados implementada, baseada nos princípios de *Big Data*, transformou dados brutos em inteligência acionável a partir do uso de visu-

alizações (painéis, gráficos, tabelas e mapas) e relatórios dinâmicos conectados ao banco de dados, identificando padrões de contaminação em áreas críticas, que contribuiu para subsidiar o Plano Nacional de Monitoramento da Qualidade Hídrica. Essa iniciativa representa um marco do fortalecimento da governança ambiental no país e viabiliza a replicabilidade e expansão, uma vez que o modelo proposto pode ser adaptado a diferentes contextos e à continuidade do monitoramento estabelecido no Plano.

Palavras-chave: *Big Data*. Plano de Monitoramento Ambiental. Gerenciamento de Dados. Qualidade da Água. Governança Ambiental.

Abstract

National environmental monitoring programs face significant challenges in consolidating and managing large volumes of heterogeneous data. In partnership with the local environmental regulatory authority, this project consists of the development of

1. Water Services and Technology (WST), São Paulo, Brasil.

2. Water Services and Technology (WST), Florianópolis, Brasil.

3. Water Services and Technology (WST), Ontário, Canadá.

a robust data infrastructure to support the formulation of the National Water Quality Monitoring Plan for a specific region in the Middle East, based on a diagnosis of water quality and on the mapping of priority areas for the investigation of water contamination. The methodology combined the structuring and validation of an extensive historical database, using data engineering methods and processes, along with the execution of a new national sampling campaign designed to fill the spatial and temporal gaps identified. The implemented data architecture, based on Big Data principles, transformed raw data into actionable intelligence using dynamic visualizations (dashboards, charts, tables, and maps) and reports connected to the database, enabling the identification of contamination patterns in critical areas and supporting the National Water Quality Monitoring Plan. This initiative represents a milestone in strengthening environmental governance in the country and enables scalability and replicability, as the proposed model can be adapted to different contexts and to the continued monitoring established by the Plan.

Key words: *Big Data. Environmental Monitoring Plan. Data Management. Water Quality. Environmental Governance.*

1. Introdução

A gestão de dados constitui um pilar estratégico para o sucesso de programas de monitoramento ambiental, especialmente em países de grande extensão territorial e elevada heterogeneidade hidrogeológica. A capacidade de coletar, analisar e interpretar informações de qualidade é o que transforma o monitoramento em uma ferra-

menta eficaz para a proteção de recursos naturais e a formulação de políticas públicas. Essa abordagem está alinhada aos princípios apresentados por Bartram e Ballance (1996), que enfatizam a importância de que programas de monitoramento gerem dados úteis e confiáveis para subsidiar a tomada de decisão e a gestão dos recursos hídricos.

A eficácia desses programas, no entanto, é frequentemente limitada pela dificuldade de consolidar e gerenciar grandes volumes de dados provenientes de múltiplas fontes, formatos e períodos históricos. A ausência de infraestruturas adequadas compromete a integração, a rastreabilidade e o aproveitamento analítico dessas informações, fazendo com que dados ambientais valiosos permaneçam dispersos e subutilizados. Nesse contexto, a crescente diversidade, complexidade e escala dos dados ambientais tem impulsionado a adoção de abordagens baseadas em *Big Data* e análises avançadas como estratégia para viabilizar o monitoramento ambiental em larga escala.

O termo *Big Data* é empregado neste estudo para caracterizar conjuntos de dados ambientais que, devido ao seu volume, diversidade, complexidade e crescimento contínuo, excedem a capacidade dos sistemas tradicionais de armazenamento e processamento e exigem infraestruturas computacionais e métodos analíticos avançados para que possam gerar informação útil para a gestão ambiental.

Essa conceituação está alinhada à definição adotada por Vance *et al.* (2024), que descrevem *Big Earth Data* como volumes massivos de dados heterogêneos e continuamente acumulados, oriundos de múltiplas fontes, cuja exploração requer ferramentas computacionais e analíticas avançadas

para a extração de conhecimento sobre o sistema terrestre. De forma complementar, o DAMA-DM-BOK (2017) caracteriza *Big Data* como dados tão grandes, rápidos ou complexos que não podem ser processados por aplicações tradicionais de gerenciamento e análise de dados, exigindo arquiteturas, tecnologias e práticas específicas de governança da informação.

Embora abordagens baseadas em *Big Data* forneçam a infraestrutura necessária para armazenar e organizar grandes volumes de dados ambientais, sua efetiva contribuição para a gestão ambiental depende da capacidade de transformar esses dados em informação interpretável e acionável. Essa transformação ocorre por meio de processos analíticos estruturados, que permitem explorar, correlacionar e interpretar os dados de forma sistemática.

Nesse cenário, *data analytics* desempenha papel central ao viabilizar o tratamento e a exploração desses grandes volumes de dados. De acordo com Ncube *et al.* (2024), *data analytics* compreende a coleta, limpeza, análise e interpretação de grandes conjuntos de dados, permitindo a extração de insights relevantes para desafios ambientais. Os autores destacam ainda que sua aplicação em programas ambientais possibilita o processamento eficiente dos dados, a identificação de padrões ocultos e a modelagem preditiva, resultando em decisões mais bem fundamentadas e no fortalecimento da governança ambiental em escala regional e nacional.

Este projeto foi desenvolvido em parceria com o órgão regulador ambiental de uma região do Oriente Médio com mais de 2 milhões de km² de extensão territorial, caracterizada por clima árido, recursos hídricos limitados e crescente de-

manda agrícola. A alta variabilidade hidrogeológica, somada a décadas de coleta dispersa de dados por diferentes instituições, resultou em um acervo fragmentado e de baixa confiabilidade. A ausência de diagnóstico nacional consolidado impedia o planejamento estratégico para conservação dos recursos hídricos no país e a identificação sistemática de áreas passíveis de contaminação hídrica, que denominaremos, neste trabalho, como “áreas críticas”.

O projeto estruturou-se com dois objetivos centrais. O primeiro consistiu em criar uma infraestrutura de dados confiável, centralizada e auditável para diagnosticar a qualidade das águas superficiais e subterrâneas e identificar possíveis áreas críticas em todo o território. O segundo objetivo foi utilizar a inteligência gerada para embasar o Plano Nacional de Monitoramento de Qualidade Hídrica, orientando decisões regulatórias e o planejamento de longo prazo para uso sustentável dos recursos hídricos da região.

Este trabalho documenta como foram estruturados os fluxos de integração, validação, rastreabilidade e análise dos dados ambientais que permitiram transformar um acervo fragmentado em uma base nacional confiável para suporte à decisão regulatória, oferecendo um modelo de referência para programas de monitoramento ambiental em larga escala.

2. Metodologia e Desenvolvimento do Projeto

A superação dos desafios de volume, variedade e veracidade dos dados exigiu uma abordagem multifacetada. A metodologia integrou

a recuperação e a validação de um vasto acervo histórico com a execução de nova campanha de amostragem em escala nacional. Toda a operação foi sustentada por uma robusta arquitetura de *Big Data*, projetada para processar, estruturar e analisar informações ambientais, garantindo a integridade e rastreabilidade necessárias para tomada de decisões estratégicas.

2. 1. Consolidação da Base Histórica (1953-2023)

O primeiro passo consistiu na compilação e organização de dados históricos provenientes de mais de 14 fontes distintas, abrangendo um período de 70 anos. O acervo inicial de qualidade hídrica reunia aproximadamente 24 mil amostras e 3 milhões de resultados analíticos, oriundos de diferentes campanhas e instituições, com metodologias e formatos variados.

A etapa de consolidação envolveu padronização de nomenclaturas, unidades e parâmetros, além de verificação de consistência espacial e temporal dos registros. Foram aplicados filtros de qualidade e validações cruzadas de metadados para identificar duplicidades, lacunas e inconsistências. Somente dados que atenderam critérios de integridade e rastreabilidade definidos no protocolo de QA/QC foram incorporados à base final.

Como resultado, cerca de 9 mil amostras e 1,4 milhão de resultados analíticos foram aproveitados, formando acervo confiável e estruturado para análises estatísticas e temporais em escala nacional. Essa base consolidada tornou-se ponto de partida para etapas seguintes de modelagem e visualização, garantindo coerência técnica e compa-

rabibilidade entre dados históricos e novos resultados de campo, seguindo os princípios metodológicos para o monitoramento e avaliação da qualidade da água (BARTRAM e BALLANCE, 1996).

2. 2. Nova Campanha Nacional de Monitoramento (2024-2025)

Com base na consolidação histórica de dados, foi planejada e executada nova campanha nacional de monitoramento entre 2024 e 2025. Esta etapa objetivou preencher lacunas espaciais, temporais e qualitativas identificadas no diagnóstico anterior, assegurando a representatividade geográfica e analítica necessária para subsidiar o Plano Nacional e a identificação de áreas prioritárias para investigação de contaminação.

A campanha abrangeu as 13 províncias do país, com mais de 2 mil novos pontos de coleta definidos a partir de análises de lacunas (*gap analysis*) e critérios técnicos de representatividade hídrica, hidrogeológica e de uso do solo. A análise de lacunas considerou densidade amostral, intervalos sem atualização e ausência de parâmetros analíticos relevantes para diagnóstico nacional.

Foram integradas informações cartográficas de bacias hidrográficas, unidades hidrogeológicas, tipos de aquífero e classes de uso e cobertura do solo, permitindo identificação de áreas sub-representadas nos dados históricos compilados anteriormente para o diagnóstico da qualidade hídrica do país. O cruzamento dessas informações orientou a seleção dos novos pontos de coleta, assegurando cobertura de representação adequada para ambientes superficiais e subterrâneos e equilíbrio de amostragens entre áreas urbanas, agrícolas e

industriais, tipos de uso frequentemente associados a processos de contaminação hídrica.

Além da componente espacial, a *gap analysis* incluiu uma avaliação temporal, que verificou intervalos entre campanhas anteriores, datas de última amostragem e séries históricas com descon-tinuidade. Essa etapa possibilitou priorizar regiões com lacunas temporais superior a cinco anos e ampliar a representatividade sazonal das medições. O processo foi desenvolvido em conjunto com especialistas locais e validado tecnicamente antes da definição do plano de amostragem, garantindo co-erência hidrogeológica, viabilidade logística e com-patibilidade com os objetivos de representatividade nacional do programa.

Foram amostradas águas subterrâneas e su-perficiais, contemplando parâmetros físicos, quí-micos e biológicos relevantes para o diagnóstico da qualidade hídrica. Para garantir padronização e rastreabilidade, adotaram-se fichas digitais de campo utilizando aplicativo WST Fielder, que per-mite o registro direto com georreferenciamento e sincronização com banco de dados central. Essa ferramenta assegurou consistência metodológica, reduziu erros de transcrição e otimizou integração entre campo e gestão de dados.

A etapa de garantia e controle de qualidade (QA/QC) foi conduzida de forma sistemática ao longo de toda a campanha, incluindo amostras de controle, duplicatas e brancos de campo e de transporte. As análises foram realizadas em labo-ratórios acreditados, operando sob a norma ISO/IEC 17025, o que assegurou a confiabilidade e rastreabilidade dos resultados.

A execução da campanha apresentou desafios

logísticos expressivos, considerando a extensão terri-torial, as condições áridas de determinadas regiões, as permissões de acesso e a variabilidade da infra-estrutura local. Houve necessidade de coordenação contínua entre as equipes de campo e os laborató-rios responsáveis pelas análises, garantindo o correto acondicionamento, transporte e processamento das amostras dentro dos prazos estabelecidos.

Durante a execução, destacou-se a impor-tância da capacitação das equipes locais quanto ao uso das fichas digitais. As atividades de trei-namento incluíram orientações sobre o preenchi-mento digital, checagem de consistência dos da-dos, inclusão de novos pontos e sincronização das informações com o sistema central. Esse processo contribuiu significativamente para elevar a quali-dade e a completude dos registros, fortalecendo a integração entre a coleta de campo e a gestão de dados.

Como resultado, a campanha produziu mais de 10 mil novos resultados analíticos provenientes de aproximadamente 2 mil amostras coletadas em todo o território nacional. A execução dessa etapa evidenciou a eficiência do uso de tecnologias digitais associadas a processos de gestão integrada de dados ambientais, reforçando a confiabilidade e a rastreabi-lidade das informações geradas em escala nacional.

2. 3. Arquitetura da Solução de Dados

Para gerenciar o grande volume, variedade e complexidade das informações provenientes tanto da base histórica quanto da nova campanha na-cional de monitoramento, foi desenvolvida uma arquitetura de solução de dados estruturada em camadas funcionais (Figura 1), garantindo gover-

nança, escalabilidade e acessibilidade ao longo do ciclo de vida da informação. Essa abordagem de engenharia de dados em multicamadas tem sido reconhecida como essencial para análises ambientais em larga escala (SNIDER et al., 2018).

A camada de ingestão coletou e consolidou dados de múltiplas fontes e formatos, armazenando-os em um *Data Lake* centralizado. Desenvolveram-se processos automatizados de ETL (*Extract, Transform, Load*) para tratar a heterogeneidade das origens, incluindo planilhas, relatórios em PDF, bancos de dados legados e os dados oriundos da nova campanha de amostragem. Essa etapa assegurou integração contínua de informações e ma-

nutenção de metadados descrevendo proveniência, formato e situação de cada conjunto de dados. A automação dos processos ETL mostrou-se fundamental para lidar com a heterogeneidade das fontes, reduzindo significativamente o tempo de processamento e minimizando erros de transcrição.

Na camada de processamento e qualidade, dados ingeridos foram submetidos a procedimentos de validação, padronização e enriquecimento, utilizando *Databricks* e *Python*. Implementou-se uma camada de qualidade (*Data Quality Layer*), responsável por identificar e corrigir inconsistências, aplicar regras de conformidade e consolidar dicionários de dados. Essa abordagem garantiu

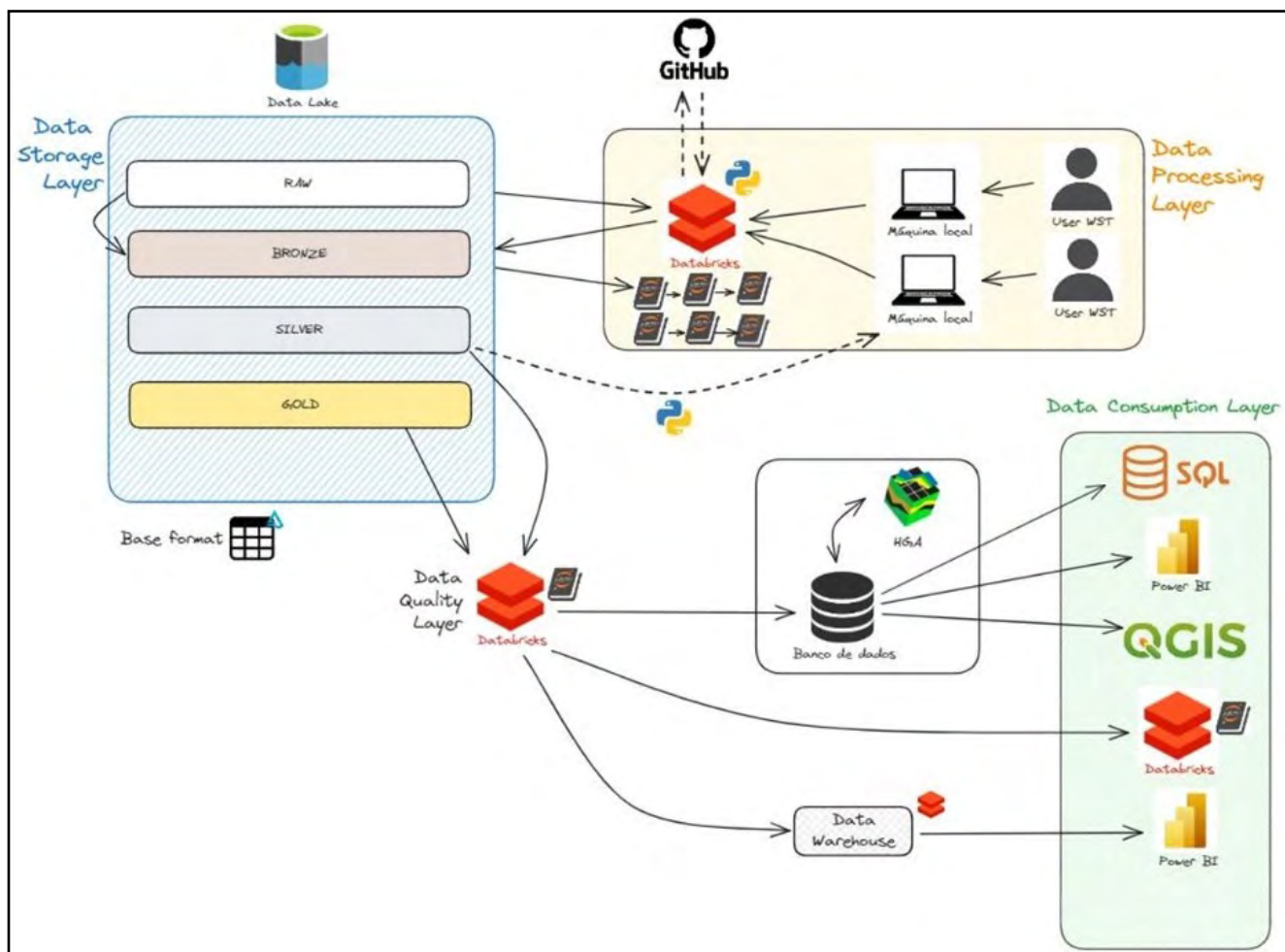


Figura 1 - Arquitetura da solução de dados estruturada em camadas de armazenamento, processamento, qualidade e consumo. Fonte: Water Services and Technologies (2024).

consistência semântica entre diferentes origens, elevando significativamente a confiabilidade da base consolidada e preparando-a para análises estatísticas, espaciais e temporais subsequentes.

A camada de consumo disponibilizou dados processados em diferentes formatos e interfaces, conforme as necessidades analíticas e operacionais do projeto. A arquitetura viabilizou uso de ferramentas de geoprocessamento (*QGIS*) para visualização georreferenciada; painéis dinâmicos (*Power BI*) para análise interativa e monitoramento de indicadores de qualidade; e consultas diretas via SQL e *software Hydro GeoAnalyst* (HGA), permitindo análises especializadas e integração com outras plataformas corporativas.

Essa estrutura modular proporcionou base sólida para gestão integrada de dados ambientais, permitindo controle de versões, rastreabilidade completa e reuso seguro dos dados em diferentes contextos analíticos. A implementação bem-sucedida da arquitetura resultou em sistema robusto capaz de integrar dados históricos fragmentados com novos registros digitalizados, mantendo rastreabilidade completa e permitindo auditorias sistemáticas de qualidade. A arquitetura desenvolveu reforçou princípios de governança digital como transparência, auditabilidade e interoperabilidade, essenciais para análises de larga escala e tomadas de decisão baseadas em evidências.

2. 4. Análise de dados e Geração de Inteligência

A base de dados consolidada, reunindo acervo histórico e resultados da nova campanha, totalizou aproximadamente 26 mil amostras e mais de 2 mi-

lhões de resultados brutos. Após aplicação de procedimentos sistemáticos de validação e depuração voltados à verificação de consistência, remoção de duplicidades e eliminação de registros fora de escopo, a base aproveitável foi composta por cerca de 11 mil amostras e 1,5 milhão de resultados validados.

Para contextualizar resultados analíticos e apoiar a definição de diretrizes comparativas, realizou-se um levantamento de padrões de qualidade da água reconhecidos internacionalmente. Esses valores de referência foram utilizados como base para a análise comparativa e interpretação dos resultados, permitindo identificação de áreas com concentrações acima dos limites estabelecidos.

A abordagem analítica concentrou-se em 46 parâmetros-chave, definidos a partir de critérios de frequência de amostragem, relevância regulatória e criticidade ambiental. Além de análises estatísticas descritivas e de tendência, conduziu-se uma avaliação geoespacial, correlacionando resultados com mapas de uso e ocupação do solo, classes hidrogeológicas e tipologias de aquífero. Essa combinação permitiu identificar padrões espaciais e temporais de contaminação, bem como potenciais relações de causa e efeito entre fontes antrópicas e respostas ambientais.

Esses resultados foram convertidos em inteligência ambiental aplicada, mediante elaboração de produtos analíticos e visuais que ampliaram a compreensão do cenário hídrico nacional. Foram desenvolvidos mapas temáticos representando distribuição espacial de contaminantes, análises estatísticas agregadas para apoio à priorização de áreas críticas e painéis interativos (*dashboards*) permitindo exploração dinâmica das informações por

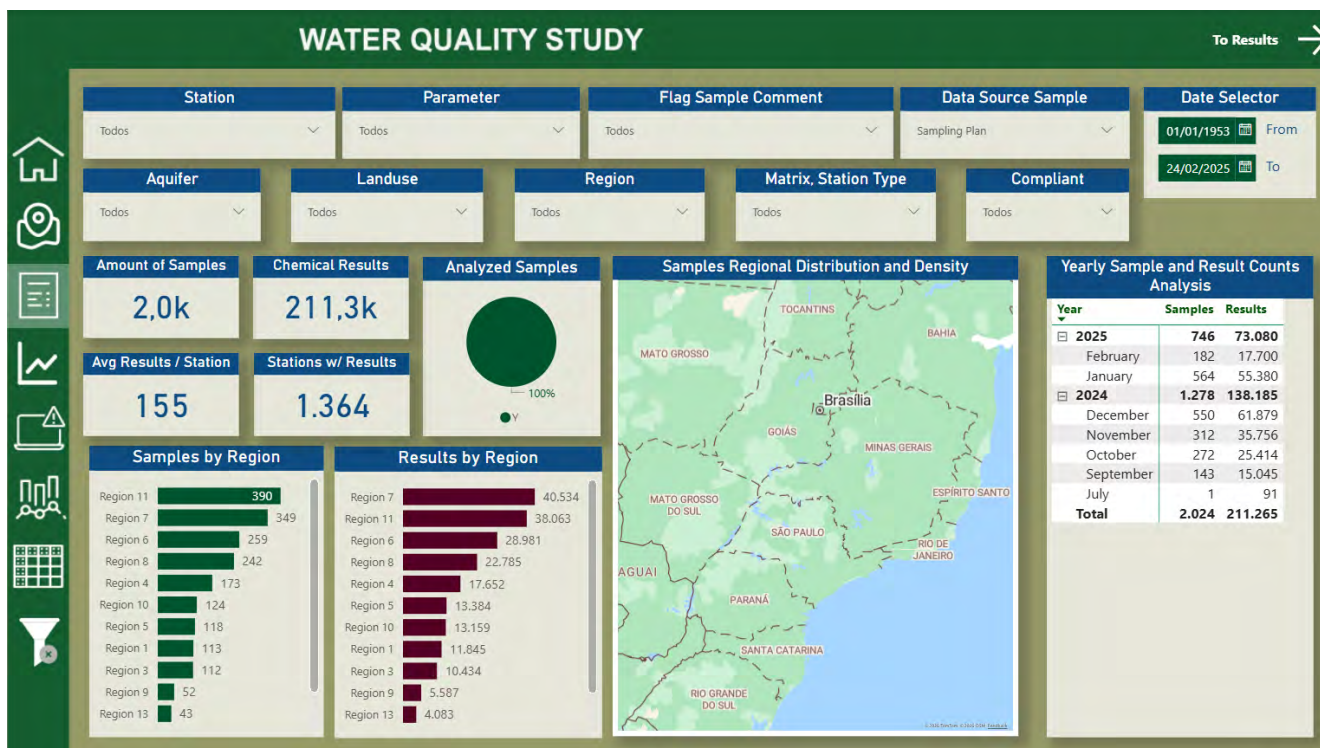


Figura 2 - Dashboard interativo desenvolvido para apresentação dos dados de qualidade da água¹. Fonte: Water Services and Technologies (2024).

gestores e tomadores de decisão (Figura 2).

A identificação sistemática dessas áreas com diferentes tipologias de contaminação subsidiou diretamente a definição de critérios para priorização de investigações detalhadas e orientou a alocação de recursos para futuras ações de remediação e acompanhamento no Plano Nacional de Monitoramento de Qualidade Hídrica.¹

2. 5. Identificação e Priorização de Áreas Críticas

O sistema desenvolvido viabilizou a identificação sistemática de áreas com diferentes perfis de contaminação, tanto na água subterrânea, quanto superficial, permitindo classificação e priorização segundo critérios técnicos e regulatórios. A integração de análises estatísticas, geoespaciais

e temporais possibilitou não apenas identificar as áreas críticas, a partir da comparação de resultados com seus limites estabelecidos em padrões de qualidade, mas também compreender padrões de evolução temporal e relações com o mapeamento de uso e ocupação do solo.

Entre os principais achados que subsidiaram a identificação de áreas prioritárias para investigação, destacaram-se:

- **Contaminação por Nitrato em Áreas Agrícolas:** Concentrações elevadas de nitrato foram identificadas sistematicamente em aquíferos situados em regiões de agricultura intensiva, com valores superiores aos padrões de potabilidade em diversos pontos. A correlação espacial com áreas de aplicação intensiva de fertilizantes nitrogenados sugere contribuição

1. Mapa deslocado para garantia da confidencialidade do projeto

antrópica significativa, caracterizando áreas prioritárias para investigação detalhada e possível implementação de medidas de remediação.

- **Intrusão Salina em Zonas Costeiras:** Presença de concentrações elevadas de sais dissolvidos foi observada em aquíferos de zonas costeiras, com padrões sugestivos de intrusão salina natural ou induzida por possível superexploração. A identificação dessas áreas é importante para gestão de recursos hídricos e pode indicar necessidade de restrições de uso ou implementação de barreiras hidráulicas.

- **Contaminação Microbiológica em Áreas Urbanas:** Tendências sazonais em parâmetros microbiológicos, especialmente coliformes totais e termotolerantes, foram observadas em corpos hídricos urbanos. A persistência de contaminação microbiológica em determinadas áreas sugere fontes contínuas associadas a deficiências em sistemas de saneamento, caracterizando-as como áreas impactadas por contaminação difusa, que requerem intervenção.

- **Contaminação por Metais Pesados em Áreas Industriais:** Foram identificadas concentrações pontuais elevadas de metais e substâncias inorgânicas em regiões com histórico de atividades industriais ou mineração, indicando necessidade de avaliação de risco e possível classificação como áreas contaminadas ou impactadas por contaminação difusa, conforme critérios regulatórios locais.

2. 6. Embasamento do Plano Nacional

A inteligência gerada pelo sistema constituiu base técnica fundamental para a estruturação do

Plano Nacional de Monitoramento de Qualidade Hídrica. O plano incorporou uma rede otimizada de pontos de monitoramento contínuo, definida com base em análises de lacunas e áreas prioritárias, bem como protocolos padronizados de coleta e análise, garantindo a comparabilidade temporal dos resultados. As frequências de amostragem foram ajustadas conforme a criticidade de cada área, assegurando maior eficiência e representatividade do monitoramento.

Além disso, foram estabelecidas diretrizes para a investigação detalhada de áreas impactadas por fontes de contaminação, orientando a priorização de ações corretivas e preventivas. A adoção de ferramentas digitais de campo e de processos automatizados de gestão de dados foi consolidada como padrão operacional, garantindo a continuidade, a escalabilidade e a sustentabilidade do sistema de monitoramento em longo prazo.

2. 7. Caracterização como Sistema *Big Data*

O projeto enquadra-se no modelo de *Big Data* pela presença marcante dos pilares fundamentais conhecidos como “4Vs”: volume, variedade, veracidade e valor (NYIKANA e IYAMU, 2023).

O volume manifesta-se pela escala nacional do projeto, amplitude da série histórica (70 anos) e número expressivo de registros (26 mil amostras, 2 milhões de resultados validados). A variedade decorre das múltiplas origens, tipos de arquivos, matrizes ambientais e estruturas de dados integradas. A veracidade foi assegurada mediante filtros de qualidade rigorosos, validação cruzada, enriquecimento de metadados e análise crítica de séries temporais. O valor materializa-se na transformação

de dados brutos em inteligência acionável para identificação de áreas críticas e embasamento de políticas públicas.

Esse perfil alinha-se com abordagens recentes de monitoramento ambiental orientado por *Big Data*, como demonstrado por Mohialden *et al.* (2024), que evidenciam o potencial de modelos preditivos aplicados a conjuntos extensos e heterogêneos de dados de qualidade hídrica, mesmo sem fluxo contínuo de informações.

2. 8. Desafios e Lições Aprendidas

Durante execução do projeto, foram superados desafios técnicos e operacionais significativos. A diversidade de formatos dos dados históricos exigiu adaptação e desenvolvimento de múltiplos processos ETL para garantir padronização sem perda de granularidade. No âmbito operacional, a implementação de tecnologias de coleta digital enfrentou dificuldades de adesão por parte de algumas equipes de campo, decorrentes de barreiras culturais e linguísticas, demandando esforços adicionais de treinamento e suporte técnico contínuo.

As experiências acumuladas resultaram em lições valiosas para aprimoramento de futuras iniciativas de governança ambiental baseada em dados:

- **Curadoria de Dados Legados:** Validação rigorosa e enriquecimento de metadados de acervos históricos mostraram-se etapas críticas para garantir integridade de análises temporais. Investir tempo e recursos nessa fase inicial é essencial para assegurar confiabilidade dos resultados e evitar vieses analíticos.
- **Flexibilidade Tecnológica:** Utilização de ferramentas flexíveis foi essencial para garantir

interoperabilidade entre diferentes plataformas e sistemas, facilitando integração de dados e colaboração entre instituições.

- **Engajamento Institucional:** Envolvimento de múltiplas instituições desde fase de planejamento foi crucial para assegurar aderência às necessidades locais, facilitar acesso a dados históricos e promover apropriação dos resultados por parte dos órgãos reguladores.

2. 9. Replicabilidade e Potencial de Expansão

A estrutura metodológica e a arquitetura de dados demonstraram não apenas eficácia, mas também alta replicabilidade e escalabilidade. O modelo proposto pode ser adaptado a diferentes contextos, sejam estes nacionais ou regionais, fortalecendo a governança de recursos naturais e promovendo tomada de decisão mais informada e transparente.

Um outro potencial de expansão identificado é a possibilidade de evolução do sistema para uma nova fase, na qual o banco de dados utilizado no diagnóstico inicial seja consolidado como repositório unificado e permanente dos dados gerados ao longo do Plano Nacional de Monitoramento de Qualidade Hídrica proposto no projeto. Essa estrutura permitiria a entrada contínua de informações pelas múltiplas instituições responsáveis pelo monitoramento, mediante protocolos padronizados de ingestão e validação. A centralização dos dados em um único ambiente auditável e interoperável fortaleceria a governança ambiental, ampliando a capacidade e tempestividade de análise temporal, rastreabilidade e integração entre diferentes ciclos de monitoramento.

Ademais, o sistema implementado possui potencial de expansão para incorporar módulos de análise preditiva utilizando técnicas de *machine learning* para antecipação de tendências de contaminação; integração em tempo real com sensores automatizados; desenvolvimento de modelos de dispersão de contaminantes; e implementação de portal *web* público para democratização do acesso às informações ambientais.

3. Conclusão

O projeto demonstrou importância estratégica da integração de *Big Data* e governança ambiental para gestão de recursos hídricos e identificação sistemática de áreas críticas no âmbito da qualidade hídrica. A combinação de coleta digital, modelagem avançada de banco de dados e análises multidimensionais possibilitou transformar um vasto e disperso conjunto de dados brutos em inteligência técnica acionável, fortalecendo as bases para decisões regulatórias e estratégicas.

Ao consolidar informações de sete décadas e complementá-las com nova campanha nacional executada com suporte tecnológico integrado, foi possível gerar diagnóstico abrangente e confiável da qualidade hídrica em escala nacional. A base de dados consolidada totalizou aproximadamente 11 mil amostras validadas e 1,5 milhão de resultados analíticos, representando redução significativa em relação aos 26 mil registros brutos iniciais, evidenciando a importância dos procedimentos de validação e controle de qualidade implementados.

A identificação sistemática de áreas críticas revelou padrões relevantes que subsidiaram diretamente o Plano Nacional de Monitoramento de

Qualidade Hídrica. Destacam-se a contaminação por nitrato em aquíferos de regiões agrícolas, com concentrações superiores aos padrões de potabilidade; a intrusão salina em zonas costeiras, indicando necessidade de gestão rigorosa da exploração de águas subterrâneas; a contaminação microbológica persistente em áreas urbanas, associada a deficiências em sistemas de saneamento; e a ocorrência de metais pesados em zonas destinadas à atividade industrial e mineração; que demandam rigoroso controle ambiental. Essa priorização permitiu otimizar a alocação de recursos para investigações detalhadas e ações de mitigação direcionadas, fortalecendo a capacidade regulatória do país.

O Plano Nacional incorporou os resultados do projeto mediante a definição de uma rede otimizada de pontos de monitoramento, protocolos padronizados de coleta e análise, e frequências de amostragem ajustadas conforme a criticidade de cada área. A adoção de ferramentas digitais de campo e de processos automatizados de gestão de dados foi consolidada como padrão operacional, garantindo a continuidade, a escalabilidade e a sustentabilidade do sistema de monitoramento em longo prazo.

O projeto representa um marco na gestão ambiental baseada em dados no Oriente Médio. Ao gerenciar elevado volume e diversidade de dados ambientais e, sobretudo, assegurar sua veracidade mediante validação rigorosa, a iniciativa gerou valor estratégico inestimável. Serve como exemplo concreto de como arquiteturas de *Big Data* podem redefinir planejamento, regulação e proteção de recursos hídricos, especialmente no contexto de identificação e controle de contaminação hídrica.

Como desdobramento, está previsto o desenvolvimento de um portal *web* público conectado ao banco de dados unificado e interoperável, que assegurará a continuidade do monitoramento e a democratização do acesso às informações. O sistema possui potencial de expansão para incorporar módulos de análise preditiva utilizando técnicas de *machine learning*; integração com sensores automatizados para monitoramento em tempo real; e desenvolvimento de modelos de dispersão de contaminantes. Essas perspectivas consolidarão a governança ambiental digital como ferramenta essencial à gestão e à proteção dos recursos hídricos em longo prazo.

3. 1. Referências Bibliográficas

BARTRAM, J.; BALLANCE, R. (Ed.). **Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programmes**. London: E & FN Spon, 1996.

DAMA INTERNATIONAL. **The DAMA guide to the data management body of knowledge (DAMA-DMBOK)**. 2. ed. Bradley Beach, NJ: Technics Publications, 2017.

EUROSTAT. **Glossary: Data analytics**. Statistics Explained. Disponível em: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary%3AData_analytics. Acesso em: 27 out. 2025.

MOHIALDEN, Y.; HUSSEIN, N. M.; SALMAN, S. **Automated water quality assessment using big data analytics**. Mesop. J. Big Data, v. 2024, p. 162-179, 2024. Disponível em: <https://doi.org/10.58496/mjbd/2024/015>. Acesso em: 27 out. 2025.

NCUBE, M.; ADEBAYO, B.; HASSAN, I. **Enhancing environmental decision-making: a systematic review of data analytics applications in monitoring and management**. Environmental Systems Research, v. 13, n. 24, 2024. Disponível em: <https://doi.org/10.1007/s43621-024-00510-0>. Acesso em: 27 out. 2025.

NYIKANA, W.; IYAMU, T. **The logical differentiation between small data and big data**. South African Journal of Information Management, v. 25, n. 1, a1701, 2023. Disponível em: <https://doi.org/10.4102/sajim.v25i1.1701>. Acesso em: 15 out. 2025.

SNIDER, D.; MORGAN, J. D.; SCHWARTZ, M.; ADKINSON, A.; BAPTISTE, D. J. **An online analytical processing database for environmental water quality analytics**. Em: Proceedings of SoutheastCon 2018. St. Petersburg, USA: IEEE, 2018.

VANCE, T.; HUANG, T.; BUTLER, K. **Big data in Earth science: Emerging practice and promise**. Science, v. 383, p. 1193, 2024. Disponível em: <https://doi.org/10.1126/science.adh9607>