

# Avaliação da Qualidade dos Dados: As Seis Dimensões

## ■ Qualidade dos Dados e suas Dimensões

- A Qualidade dos Dados é definida como a adequação dos dados em questão para o propósito para o qual foram destinados - **ou seja, como o valor real do conjunto de dados.**
- Dimensão da Qualidade dos Dados é um termo reconhecido utilizado por profissionais de ciência de dados para descrever uma **característica dos dados** que pode ser medida ou avaliada em relação a padrões definidos **para determinar a qualidade dos dados.**
- A abordagem de avaliar diferentes dimensões da qualidade dos dados é reconhecida como útil para **identificar e entender deficiências potenciais nos dados em várias áreas**, incluindo a análise de dados ambientais.
  - Neste resumo, foi escolhida uma abordagem muito comum na comunidade científica, que considera 6 dimensões principais.

# ■ As Seis Dimensões da Qualidade dos Dados



## 1. Precisão

Refere-se à proximidade dos dados em relação aos valores verdadeiros.

Dados precisos são cruciais na análise ambiental, pois erros de medição ou coleta podem levar a conclusões incorretas e decisões inadequadas.

Por exemplo, em crises ambientais, como derramamentos químicos, a precisão dos dados é essencial para determinar o impacto real no ecossistema e na saúde humana.



## 2. Completude (Acessibilidade)

Refere-se à completude dos dados, além de garantir que todas as informações relevantes estejam sempre disponíveis.

Assegurar a completude significa, indiretamente, avaliar a relevância espacial e temporal dos dados, uma vez que um banco de dados com lacunas não pode ser considerado completo.

Assim, a completude permite uma rápida avaliação de quaisquer desvios e decisões assertivas com base em dados atualizados e representativos.



## 3. Consistência

Refere-se à uniformidade e harmonização dos dados em todo o conjunto de dados.

É importante que os dados coletados e relatados sigam padrões consistentes para facilitar a análise e comparação de diferentes informações.

A consistência dos dados é crítica em situações de crise, pois permite que as equipes de resposta e os tomadores de decisão tenham uma visão clara e coerente da situação.



## ■ As Seis Dimensões da Qualidade dos Dados



### 4. Confiabilidade (Integridade)

Refere-se à credibilidade e à fonte dos dados. É essencial garantir que os dados sejam coletados por profissionais treinados, seguindo os melhores procedimentos de amostragem, além de serem analisados por laboratórios acreditados que sigam as melhores práticas.

Adicionalmente, para evitar que os dados percam a confiabilidade, é importante que não sejam modificados devido a manuseio inadequado, como uma transcrição incorreta de papel para uma planilha eletrônica, por exemplo.



### 5. Unicidade

Refere-se a garantir que não haja dados duplicados nos resultados, locais de amostragem ou em qualquer outra informação. Isso previne vários tipos de erros analíticos.

Por exemplo, que dados coletados no mesmo local sejam considerados como diferentes pontos de amostragem, ou que a mesma informação seja contabilizada duas vezes na análise porque foi registrada em dois documentos diferentes, ambos inseridos no banco de dados.



### 6. Validade

Refere-se à adequação dos dados para os objetivos específicos da análise e ao cumprimento de parâmetros de qualidade pré-estabelecidos: os parâmetros corretos, nos lugares e matrizes corretos.

Além disso, os dados ambientais devem atender a requisitos mínimos para que sua análise não seja prejudicada: exemplos incluem o prazo de retenção (vital para a precisão dos dados) e a rastreabilidade através de cadeias de custódia (vital para a confiabilidade).

- ## Desafios da Qualidade dos Dados - Exemplos

  - Múltiplas fontes de dados
  - Compartilhamento frequente de fontes de dados incorretas
  - Muitos tipos diferentes de fontes de dados, especialmente planilhas e relatórios de laboratório em formatos variados
  - Inconsistências dentro da mesma fonte de dados: colunas com unidades inconsistentes ou até mesmo incorretas, erros de digitação nos dados de resultados, analitos com múltiplos nomes, informações faltando em planilhas supostamente consolidadas
  - Múltiplos nomes para o mesmo local, locais com várias coordenadas espaciais, ou locais sem coordenadas

## ■ Conclusão

- A avaliação da qualidade dos dados é essencial para garantir resultados confiáveis, precisos e significativos, especialmente em contextos onde as decisões são impulsionadas por dados, como no monitoramento ambiental ou na gestão de crises.
- As seis dimensões—**Precisão, Completude, Consistência, Confiabilidade, Unicidade e Validade**—fornecem uma estrutura organizada para detectar deficiências potenciais e manter altos padrões de dados.
- No entanto, alcançar esses padrões requer mais do que apenas compreensão teórica.

## ■ Conclusão

- A especialização das pessoas que atuam em dados desempenha um papel fundamental em:
  1. Identificar lacunas e inconsistências nos dados que podem não ser imediatamente óbvias.
  2. Garantir processos adequados de coleta de dados e conformidade com padrões regulatórios.
  3. Mitigar riscos de duplicação ou problemas de integridade ao longo do ciclo de vida dos dados.
  4. Alinhar práticas de gestão de dados com as necessidades específicas da organização ou da indústria.





## ■ Conclusão

- O envolvimento de um especialista em qualidade de dados é indispensável.
  - Sua expertise assegura que os dados sejam completos, válidos e confiáveis, permitindo que os tomadores de decisão confiem nas informações extraídas.
- Um banco de dados bem estruturado é essencial para a organização, acesso e processamento eficaz dos dados ambientais.
  - Essa estrutura permite a extração eficiente de informações relevantes, garantindo que os dados sejam plenamente utilizados para fornecer *insights* acionáveis.
- Dada a complexidade dos dados ambientais e os diversos formatos e fontes a orientação especializada é crucial para garantir a extração, integração e interpretação adequadas para uma tomada de decisão informada.