

www.inthemine.com.br

UMA PUBLICAÇÃO FACTO EDITORIAL

# in the mine

Ano XVI | 2023 | Nº105 | R\$ 25,00 GESTÃO DE PROCESSOS E TECNOLOGIA PARA MINERAÇÃO



ESPECIAL

## REJEITOS & SOLUÇÕES

COMO EVOLUIU A GESTÃO DOS REJEITOS DO BENEFICIAMENTO MINERAL, A PARTIR DE SUA GERAÇÃO OU ARMAZENAMENTO

PERSONALIDADE

INÁCIO MELO: UM EMPRESÁRIO DA CONSTRUÇÃO CIVIL COMO DIRETOR-PRESIDENTE DO SGB

LEGISLAÇÃO

DIREITO DA MINERAÇÃO: OS CONCEITOS E CLASSIFICAÇÃO DOS ATOS E FATOS JURÍDICOS

MERCADO

A CONSOLIDAÇÃO DO LÍTIO NO BRASIL: HISTÓRICO, MINAS, PROJETOS E PERSPECTIVAS

GEOLOGIA

CRITÉRIOS E PARÂMETROS PARA ANALISAR E VALIDAR A QUALIDADE DE DADOS



Por *Gláucia Cuchierato<sup>1</sup>*

# CRITÉRIOS E PARÂMETROS DE QUALIDADE DE DADOS

Nesta nova série de artigos técnicos sobre a importância da qualidade da informação utilizada na declaração de recursos e reservas minerais, resultantes da pesquisa de doutoramento de Cuchierato (2022), foram indicadas nos artigos anteriores as definições relacionadas aos dados, informação e sistemas de gestão e governança, de acordo com diversos autores.

A análise da qualidade em um sistema de informações é realizada por diversos critérios de difícil determinação, por possuírem forte componente abstrato e poderem ser influenciados por fatores externos e pessoais.

Batini et al. (2009) apresentaram uma revisão sobre as possíveis metodologias para análise, avaliação e melhoria da qualidade dos dados e apontaram os critérios e dimensões que podem ser utilizados para conduzir essa avaliação e como escolhê-los. Os tipos de dados analisáveis estão conceituados no **Quadro 1**.

Também discutiram as estratégias para aprimorar a qualidade dos dados que podem ser orientadas por:

- **data-driven (baseado em dados):** melhoram a qualidade dos dados modificando diretamente seu valor; e
- **process-driven (baseado em processos):** melhoram a qualidade dos dados redesenhando os processos que os criam ou os modificam.

**Quadro 1:** Tipos de dados

TERMO	DEFINIÇÕES
<b>Estruturados</b>	<ul style="list-style-type: none"> <li>• Agregações ou generalizações de itens descritos por atributos elementares definidos dentro de um domínio.</li> <li>• Os domínios representam o intervalo de valores que podem ser conferidos a atributos e geralmente correspondem a tipos de dados elementares de linguagens de programação, como valores numéricos ou strings de texto.</li> <li>• Exemplo: tabelas relacionais e dados estatísticos.</li> </ul>
<b>Não estruturados</b>	<ul style="list-style-type: none"> <li>• Sequência genérica de símbolos, normalmente codificados em linguagem natural.</li> <li>• Exemplos: questionário contendo texto livre respondendo a perguntas abertas ou o corpo de um e-mail.</li> </ul>
<b>Semiestruturados</b>	<ul style="list-style-type: none"> <li>• Possuem uma estrutura com algum grau de flexibilidade.</li> <li>• XML é a linguagem de marcação comumente usada para representar dados semiestruturados.</li> </ul>

Fonte: **Batini et al. (2009)**

**Figura 01:** Critérios de qualidade da informação, por DAMA (Barbosa & Lyra, 2019)

<b>PRECISÃO</b>	Grau em que os dados descrevem corretamente o objeto ou evento do "mundo real".
<b>TEMPORALIDADE</b>	Problemas de tempo (além da própria tempestividade): é estável, mas responde a solicitações de mudança legítimas?
<b>CONFIANÇA</b>	A governança, a proteção e a segurança dos dados estão vigentes? Qual é a credibilidade dos dados? É verificável?
<b>COMPLETUDE</b>	A proporção de dados armazenados contra o potencial de 100%.
<b>OPORTUNIDADE</b>	Grau em que os dados representam a realidade a partir do momento requerido.
<b>EXCLUSIVIDADE</b>	Nenhuma instância de entidade será registrada mais de uma vez após identificada.
<b>VALIDADE</b>	Dados são válidos se estiverem em conformidade com a sintaxe (formato, tipo, intervalo) de sua definição.
<b>CONSISTÊNCIA</b>	A ausência de diferença ao comparar duas ou mais representações de uma instância com uma definição.
<b>USABILIDADE</b>	Os dados são compreensíveis, relevantes, acessíveis, mantidos e no nível certo de precisão?
<b>FLEXIBILIDADE</b>	Os dados são comparáveis e compatíveis com outros dados? Possui agrupamentos e classificações úteis? Pode ser reaproveitado? É fácil de manipular?
<b>VALOR</b>	Existe um levantamento de custo/ benefício para os dados? E/Ou serão utilizados da melhor maneira possível? Põe em risco a segurança ou a privacidade das pessoas ou as responsabilidades legais da organização? Apoia ou contradiz a imagem corporativa ou a visão estratégica corporativa?

**Quadro 2:** Características de dados de qualidade (Barbosa & Lyra, 2019)

TÓPICO	CARACTERÍSTICAS
<b>Características gerais</b>	<ul style="list-style-type: none"> <li>Dados ainda não são contabilizados no balanço da maioria das organizações. É complexa a atribuição de um valor monetário aos dados e difícil mensurar como eles contribuem para o sucesso organizacional.</li> <li>São duráveis e não se desgastam; são fáceis de copiar e transportar, mas não de serem reproduzidos, se perdidos ou destruídos.</li> <li>São dinâmicos e podem ser usados para várias finalidades. Os mesmos dados podem até ser usados por várias pessoas ao mesmo tempo.</li> </ul>
<b>Valor do dado</b>	<ul style="list-style-type: none"> <li>O valor dos dados pode e deve ser expresso em termos econômicos.</li> <li>Chamar dados de um ativo implica em atribuir seu valor.</li> <li>Os dados não são consumidos quando usados, como são consumidos os ativos financeiros e físicos. Têm características únicas que os diferenciam de outros ativos e não são tangíveis: os ativos físicos, por exemplo, podem ser apontados, tocados e movimentados. Os ativos financeiros, por sua vez, são contabilizados no balanço patrimonial.</li> <li>Embora existam técnicas para medir o valor qualitativo e quantitativo dos dados, ainda não existem padrões para isso. As organizações que desejam tomar melhores decisões sobre seus dados devem desenvolver maneiras consistentes de quantificar esse valor. Também devem medir os custos de dados de baixa qualidade e os benefícios de dados de alta qualidade.</li> </ul>
<b>Riscos</b>	<ul style="list-style-type: none"> <li>Dados mal gerenciados apresentam riscos éticos e de segurança.</li> <li>Dados imprecisos, incompletos ou desatualizados indicam riscos, porque suas informações não estão corretas.</li> <li>Dados podem ser facilmente copiados e replicados, o que significa que podem ser violados sem serem “retirados” de seus legítimos proprietários.</li> <li>Ameaças: <ul style="list-style-type: none"> <li>(1) Uso indevido: se os consumidores de dados não tiverem informações suficientes e corretas (metadados) sobre os dados que usam, existe o risco de eles serem mal utilizados ou compreendidos.</li> <li>(2) Não Confiabilidade (Unreliability): se a qualidade e a confiabilidade dos dados não forem estabelecidas por meio de normas e métricas, existe risco de que decisões sejam baseadas em dados não confiáveis.</li> <li>(3) Uso inadequado: se os dados não estiverem protegidos, existe o risco de que eles sejam utilizados por pessoas não autorizadas para fins não autorizados.</li> </ul> </li> </ul>
<b>Características dos dados de baixa qualidade</b>	<ul style="list-style-type: none"> <li>Custam dinheiro</li> <li>Se não atenderem às necessidades de seus consumidores, o esforço para coletar, armazenar, proteger e permitir-lhes o acesso é desperdiçado.</li> <li>Muitos dos custos são ocultos e indiretos e, portanto, difíceis de mensurar.</li> </ul>
<b>Consequências da baixa qualidade dos dados</b>	<ul style="list-style-type: none"> <li>Retrabalho.</li> <li>Soluções alternativas e processos de correção ocultos.</li> <li>Ineficiências organizacionais ou baixa produtividade.</li> <li>Conflito organizacional.</li> <li>Insatisfação do usuário.</li> <li>Custos de oportunidade, incluindo a incapacidade de inovar.</li> <li>Custos ou multas de conformidade.</li> </ul>
<b>Como mensurar o valor dos dados</b>	<ul style="list-style-type: none"> <li>Custo de obtenção e armazenamento de dados.</li> <li>Custo da substituição de dados se forem perdidos.</li> <li>Impacto para a organização se houver perda de dados.</li> <li>Custos potenciais de riscos associados aos dados.</li> <li>Custo de mitigação de riscos.</li> <li>Custo de melhoria de dados.</li> <li>Benefícios de dados de maior qualidade.</li> <li>Vantagens de compartilhar esses dados.</li> <li>Retorno esperado pelo uso inovador desses dados.</li> </ul>

Fonte: *Batini et al. (2009)*

Barbosa & Lyra (2019) consideram os dados como um “ativo da organização” e apresentam suas diversas características, sintetizadas no **Quadro 2**.

A **Figura 1** apresenta os termos conforme definições da DAMA (The Data Management Association), organização internacional sediada no Reino Unido, que visa promover a compreensão, desenvolvimento e prática da gestão de dados e informações (BARBOSA & LYRA, 2019). Entende-se que a qualidade dos dados pode ser melhorada somente quando se corrige a causa-raiz das não-conformidades de atendimento de cada parâmetro, por meio da avaliação dos processos de maneira oportuna, econômica, eficaz e eficiente. Para tanto, faz-se necessária a capacidade de medição, gerenciamento e divulgação da qualidade dos dados.

Nas próximas edições desta série de artigos serão apresentados os fundamentos da metodologia GDQM – GeoData Quality Management, definida com este objetivo. ■

### Referências:

- BARBOSA, W.; LYRA, R. Enap. Governança de Dados: Princípios, importância e desafios do Gerenciamento de Dados. Brasília, 2019. Disponível em: <https://repositorio.enap.gov.br/handle/1/5008>
- BATINI, C.; CAPIELLO, C.; FRANCALANCI, C.; MAURINO, A. Methodologies for data quality assessment and improvement. [s.l.]: ACM Computing Surveys, v. 41, n. 3, p. 2–52. 2009.
- CUCHIERATO, G. (2022), A importância da qualidade da informação no processo de declaração de recursos minerais. 293 f. (Tese de Doutorado em Engenharia de Minas). Departamento de Engenharia de Minas e do Petróleo da Escola Politécnica da Universidade de São Paulo, São Paulo, 2022.
- ISO - INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO / TS 8000-1: Qualidade de dados - Parte 1: Visão geral. [s.l.; s.n.], 2022.

1 Geólogo e Mestre em Recursos Minerais pelo IGc-USP, Doutora em Engenharia Mineral pelo PMI-EPUSP (Projeto: “O valor da qualidade da informação no processo de declaração de recursos minerais”) e Diretora Executiva da GeoAnsa-ta Projetos e Serviços em Geologia