

Qualidade de Dados

Existe uma percepção incorreta de que a qualidade de dados esteja estritamente relacionada com os seus aspectos estruturais. A qualidade de dados deve ser analisada em diversos aspectos, dos quais, muitos não estão diretamente relacionados à forma física do dado, mas ao seu entorno (BARBIERI, 2019).

Barbieri (2019) comenta que há vários autores com variadas classificações de dimensões de qualidade de dados. Abaixo são listadas algumas das dimensões mais usualmente encontrados na literatura:

- **Compleitude:** Se falta algum atributo de dados do seu registro, como CEP no endereço, não há completude;
- **Unicidade:** Se existem dois funcionários com o mesmo CPF, por exemplo, há problemas de unicidade;
- **Razoabilidade:** Se a data da receita médica é posterior à data preenchida de compra do medicamento na farmácia, por exemplo, não há razoabilidade;
- **Integridade:** Se o mesmo dado possui valores diferentes em locais distintos, não há integridade;
- **Temporalidade:** Esta característica define a necessidade do dado estar disponível no tempo exigido e na forma demandada. Por exemplo, a marcação de um assento de voo deverá ocorrer em tempo real para que a sua disponibilidade seja garantida;
- **Validade:** Está relacionada com a forma estrutural estabelecida. Por exemplo, se o código de CEP é válido e estão listados nos dados de referência. Se o tipo de dado definido está sendo obedecido, etc.;
- **Cobertura:** Dimensão de qualidade que considera o quanto os dados disponíveis atendem, por exemplo, geograficamente, aos seus objetivos. Ex. tenho todos os dados relativos ao Brasil, mas pretendo atuar na Argentina;
- **Precisão:** Dimensão de qualidade relativa à precisão do dado cadastrado. Ex: Moro no bairro centro, mas o CEP cadastrado é da pampulha.

Com a evolução da ciência de dados, outros aspectos, como privacidade e ética, passarão a ser considerados dentro dos aspectos de qualidade.

Kimball e Ross (2013) pontuam que iniciativas puramente técnicas para solucionar problemas de qualidade de dados geralmente não funcionam, a não ser que façam parte de uma cultura geral de qualidade, estabelecida pela alta cúpula da instituição. Deste modo, Kimball e Ross (2013) afirmam que problemas com qualidade de dados não podem ser solucionados apenas pela TI. Em muitos casos, na realidade, problemas com qualidade de dados nada tem haver com TI.

Referências

Revisão #12

Criado Wed, Mar 16, 2022 10:02 AM por FLAVIO LOPES DE MORAIS

Atualizado Fri, Jul 28, 2023 9:35 AM por FLAVIO LOPES DE MORAIS