Data Quality - Notes

Nils Rechberger

2026-02-09

Lecture 01: Introduction

Relevance of Data Quality

A key advance in data quality emerged in the 1920s through R. A. Fisher's work in experimental design, which introduced randomization and replication to estimate error, bias, and precision.

Definitions

Data

Data are abstract representations of selected features of real-world entities, events, and concepts, expressed and understood through clearly definable conventions (Sebastian-Coleman, 2013).

Data Quality

- Contextual data quality: The quality of data is defined by two related factors: how well it meets the expectations of data consumers (how well it is able to serve the purposes of its intended use).
- Intrinsic data quality: How well it represents the objects, events, and concepts it is created to represent.

Data Management

DQ as part of Data Management

Data Governance sits at the center (the hub) because it provides oversight, direction, policies, and coordination. Data quality is but one element of effective data management.

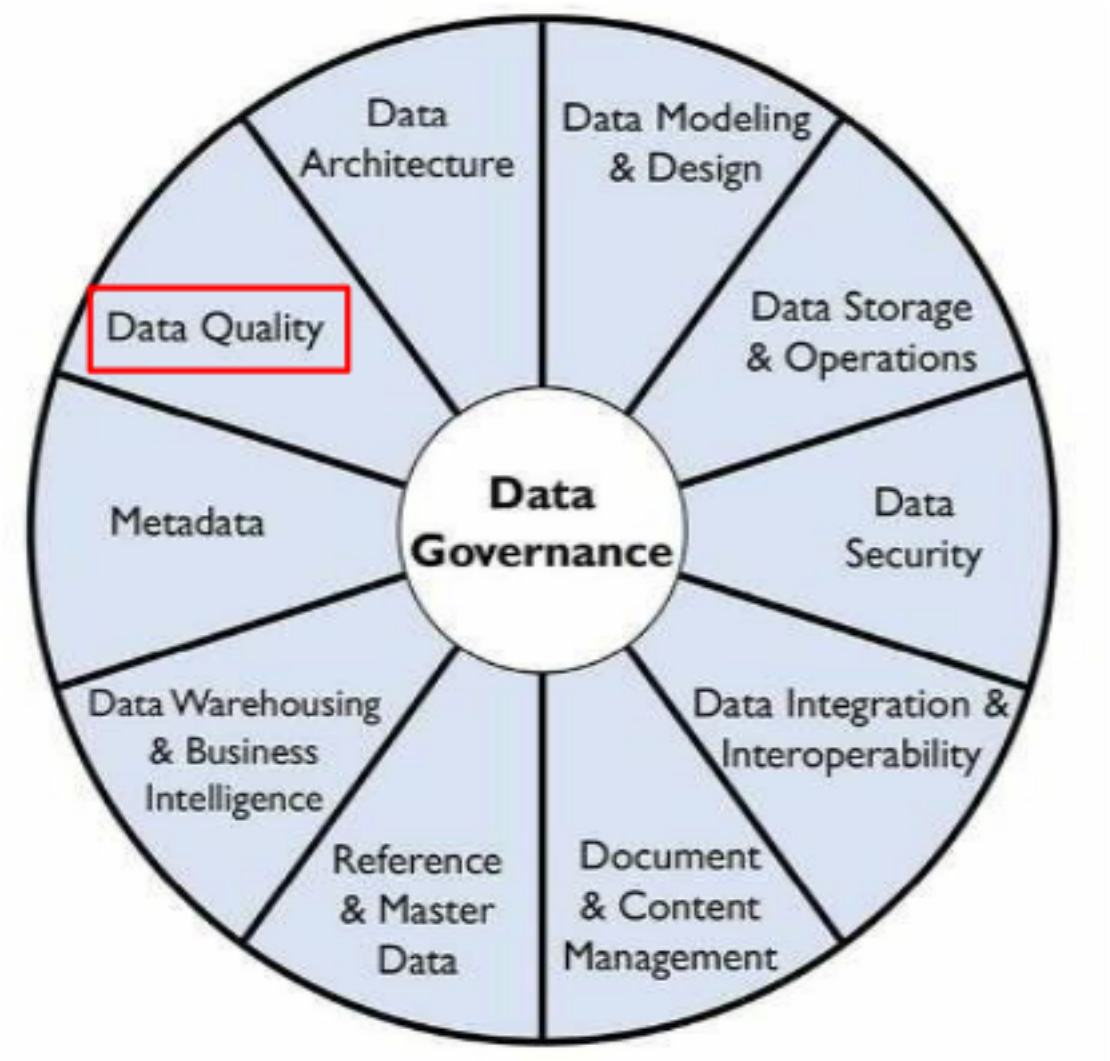

Figure 1: Data Management

Role of Data Quality Managers

- Develop a governed approach to make data fit for purpose based on data consumers requirements.
- Define standards, requirements, and specifications for data quality controls as part of the data lifecycle.
- Define and implement processes to measure, monitor, and report on data quality levels.
- Identify and advocate for opportunities to improve the quality of data, through process and system improvements.

Data Quality Dimensions

- Correctness/accuracy: The data accurately describe the entity in question
- Completeness: No missing records/field values
- Conformity/validity: Correct types, value ranges, etc.
- Consistency: No contradictions between different data sets / tables

  Note: Breaking the issue down into dimensions helps to quantify issues.

> **Note**
>
> Most sources agree that the concept of data quality has several dimensions. But they don't agree on what these dimensions are.