Data Quality Dimensions

Decisions are only as good as the data behind them. Data quality means how well data fits the purpose for which we use it. DAMA, the data management association behind the DAMA-DMBOK guide, describes data quality through dimensions, which are measurable aspects of quality. A DAMA UK working group recommended six core dimensions in 2013: accuracy, validity, completeness, uniqueness, consistency and timeliness. We can follow four parts: accuracy and validity, completeness and uniqueness, consistency and timeliness, and measuring and improving.

1. Accuracy and Validity

Accuracy asks whether data describes the real thing correctly. A customer's phone number is accurate only if it really is that customer's number. Validity asks whether data follows the rules of its format and its allowed values, such as a date in a proper format or a country from an allowed list. Data can be valid but not accurate, because a perfectly formatted phone number can still belong to someone else.

Example: A form asks for a birth date. Applying Accuracy and Validity, a date such as the thirty-first of February fails the validity check, while a real date typed with the wrong year passes it but is not accurate.

At work: Rami checks the customer list. Applying Accuracy and Validity, he finds email addresses with a broken format, which are invalid, and a phone number that belongs to a former customer, which is inaccurate. He sends both for correction.

2. Completeness and Uniqueness

Completeness asks whether the data we need is present. A record with no delivery address is incomplete for a delivery. Uniqueness asks whether each real thing appears only once, without duplicates. Duplicates lead to double counting, repeated mailings and confusion about which record is right.

Example: A party guest list has two blank names, and one guest appears twice. Applying Completeness and Uniqueness, the host fills the gaps and removes the duplicate, so the seat count is right.

At work: Rami finds that four hundred customers have no postcode, and some appear twice. Applying Completeness and Uniqueness, he asks for the missing postcodes and merges the duplicates. Mailing costs fall, because each customer receives one letter.

This document is a property of Asaseeyat Academy - © Copy Right Protected.

3. Consistency and Timeliness

Consistency asks whether the same fact agrees across different places. If the billing system says a customer lives in one city and the shop system says another, the data is inconsistent. Timeliness asks whether the data is up to date enough for its use. Last year's stock count may be fine for a yearly report, but not for deciding today's order.

Example: A person's address is changed in one app but not in another. Applying Consistency and Timeliness, the two apps disagree, and the older one is out of date.

At work: Rami compares the billing system with the shop system. Applying Consistency and Timeliness, he finds customers with two different addresses, and one system that updates only once a week. He asks for daily updates.

4. Measure and Improve

Each dimension can be checked with simple tests. For example, we can count the share of records with a missing value, the share of duplicates, or the age of the latest update. We then set a target for each dimension, based on how important the data is. When a problem appears, it is best to fix its cause at the source, such as a form field, instead of cleaning the same mistake again and again. Giving each data set a clear owner keeps the checks going.

Example: A school counts each term how many student records lack a parent's phone number. Applying Measure and Improve, the school makes that field required on the form, and the share falls.

At work: Rami measures the share of customers with a missing postcode: twelve percent. Applying Measure and Improve, he makes the postcode field required on the online form. The next month, the share has fallen.

Conclusion

In summary, data quality means fitness for use, and it can be described through dimensions. Accuracy and validity ask whether data is right and allowed, completeness and uniqueness ask whether it is whole and not repeated, and consistency and timeliness ask whether it agrees and is current. The six dimensions recommended by the DAMA UK working group are a common starting point. By measuring them and fixing causes at the source, we make decisions more trustworthy.

Sources

- DAMA UK Working Group (October 2013). The Six Primary Dimensions for Data Quality Assessment (white paper).
- DAMA International: DAMA-DMBOK, Data Management Body of Knowledge (2nd edition, 2017) – https://www.dama.org/cpages/body-of-knowledge
- UK Government Data Quality Hub: Meet the data quality dimensions – https://www.gov.uk/government/news/meet-the-data-quality-dimensions
- DAMA and DAMA-DMBOK are trademarks of DAMA International.

This document is a property of Asaseeyat Academy - © Copy Right Protected.