

Towards Trust in Health Data: Is the European Dream Achievable?

Who am I?

Mary Mangru, ex NNIT (under notice right now, starting another position after a few weeks holiday).

Email: marymangru@hotmail.com. LinkedIn: [linkedin.com/in/mary-mangru-84b75a6](https://www.linkedin.com/in/mary-mangru-84b75a6).

Many, many years experience in IT (35+), starting with Computer Science in UCD, working as a programmer, then moving onto the analysis side (business analyst - process, business rules) and then specialising in Data.

Founding president of DAMA Ireland, and on the board of the current instance.

I have a lot of experience managing data in different application areas, in many countries, though Pharma has always been to the fore, and in latter years Healthcare too.

5 minute crash course in Data Management

The Five Pillars

Data Culture: the collective behaviours, beliefs, and values within an organization that encourage employees to use data as the primary basis for everyday decision-making.

Data Operations: focused on automating and optimising the delivery, quality, and governance of the flow of data through an organisation

Data Capability: the capacity to successfully manage, secure, and extract ultimate value from its data assets through the alignment of people, processes, and technology

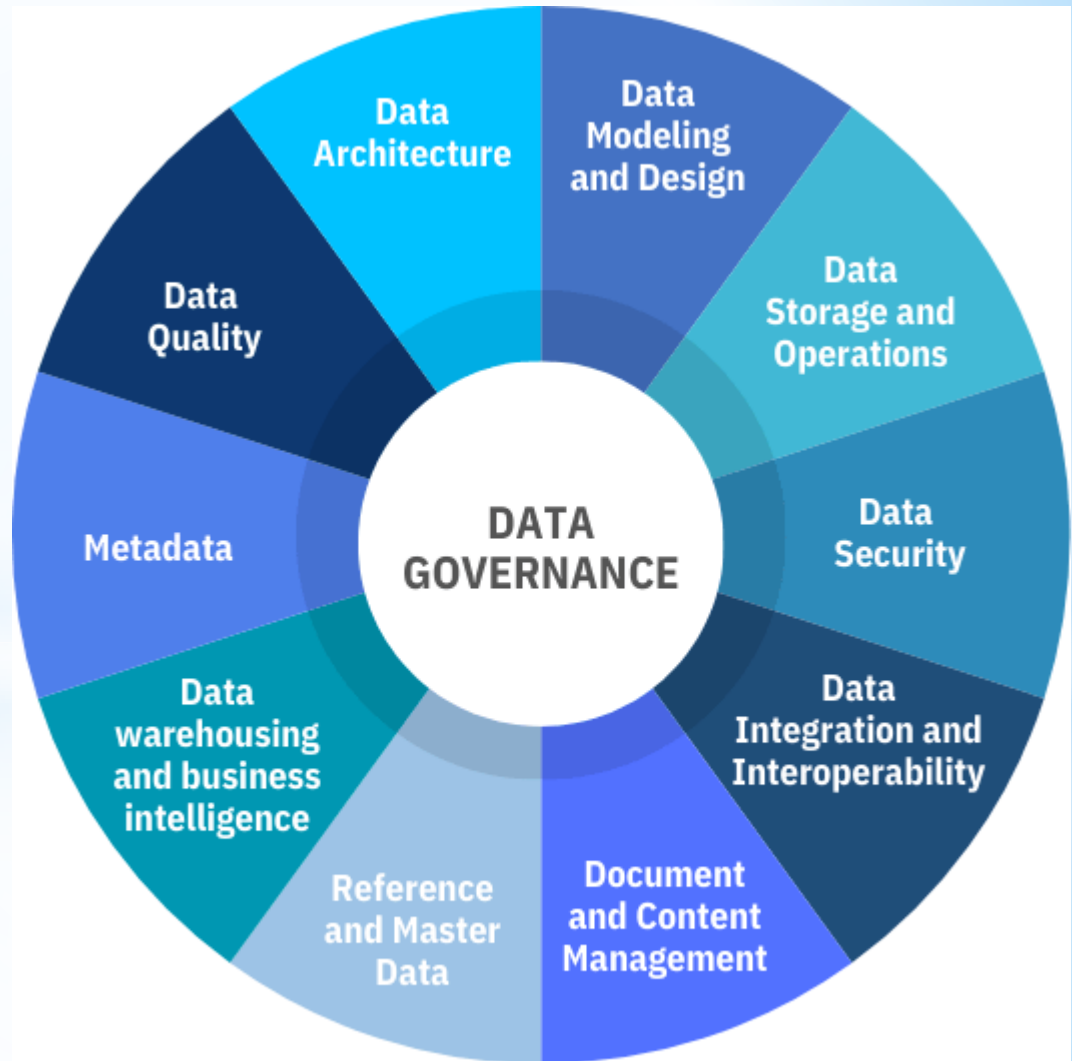
Technology: covers the physical software, hardware, cloud infrastructure, and technical tools used to collect, store, process, protect, and deliver data across an enterprise

Compliance and Ethics: covers the legal boundaries, industry regulations, and moral principles that dictate how an organization can legally and responsibly collect, store, process, and share its data

5 minute crash course in Data Management

The Data Framework -
DAMA DMBOK ©

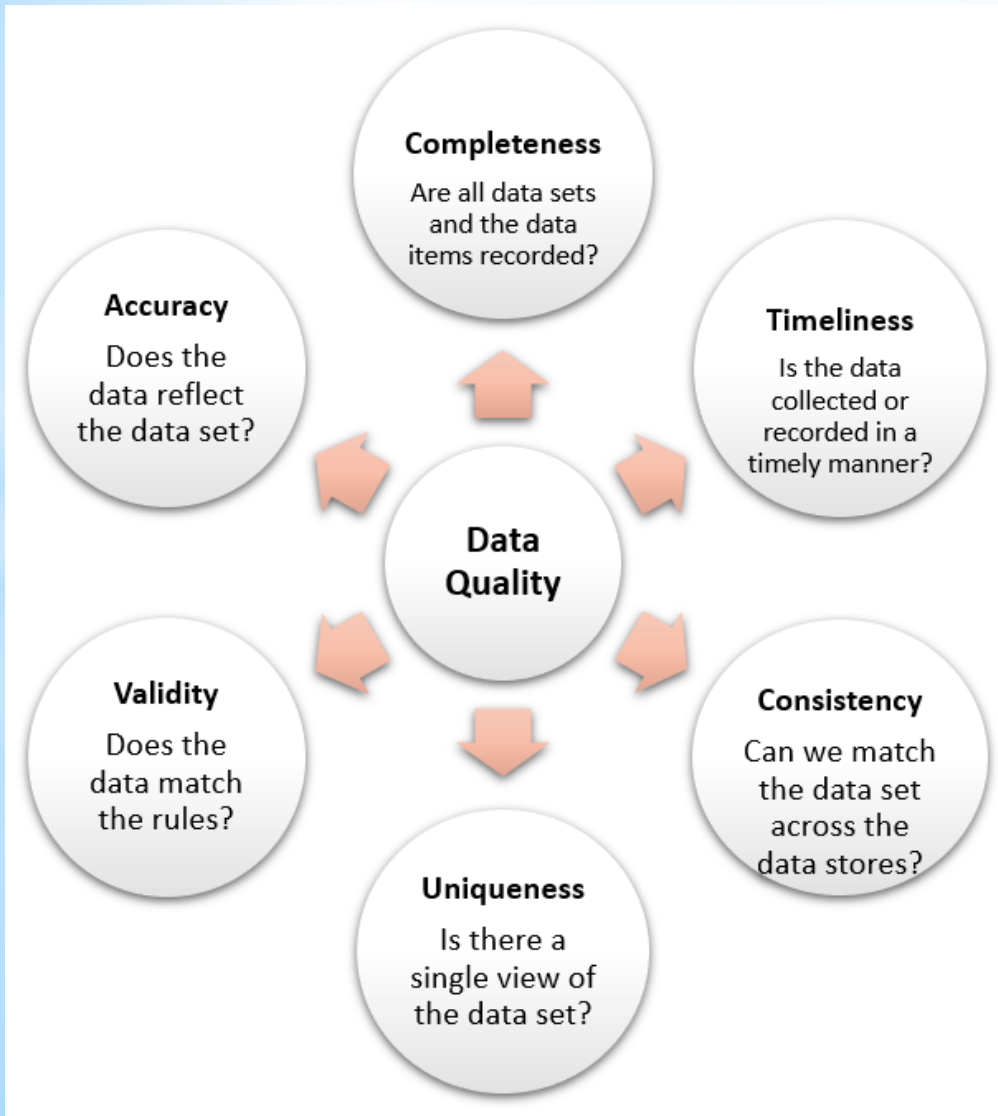
10 Segments - all ruled by
Data Governance



Copyright © 2024 DAMA International

5 minute crash course in Data Management - Data Quality

6 Common Quality Dimensions:



- Lineage
- Structure

ALCOA, ALCOA+ and ALCOA++

- Attributable
- Legible,
- Contemporaneous,
- Original
- Accurate
- Complete (+)
- Consistent (+)
- Enduring (+)
- Available (+)
- Traceable (++)

Do I need to worry about data quality? Tom Redman's the FAM Method

- * FAM provides a quick and effective data quality evaluation method that can be completed in about an hour on a Friday afternoon when the pace of work has slowed down – hence, its name. This method allows teams to assess data quality on a weekly basis and identify and address red flags before issues escalate.
- * **Assemble:** Assemble a representative sample of data from the most recent data-related activities in the department. For example – for the Sales Dept, this could be the last 100 record entries in the CRM. You can use the data that was recently created or recently used. Then select 10-15 data elements in those 100 records that are of importance to you – for instance customer names, transaction dates, amounts etc.
- * **Identify:** Review the 100 records (looking specifically at the 10-15 selected data elements). Mark all with data quality errors – for example missing or null values, invalid formats or entries, misspellings or inconsistencies. Each record is marked as either 'Perfect' or 'Defective'. This exercise will typically involve people deeply familiar with the data since there may be a small number of records that require deeper discussions on what constitutes an error for a data element.
- * **Calculate:** This is easy to do since the original number of records was 100. So if 62 records are 'Perfect' and 38 are 'Defective', the Error Rate is 38%.
- * **Quantify:** The FAM method then calculates the estimated cost of poor quality data – this can be particularly useful when presenting the concepts to C-level stakeholders. The cost calculation requires applying the **Rule of Ten*** – which states 'it costs ten times more to complete a unit of work when the data is defective, as compared to when it is perfect'.

For example, if the cost of a single unit of work is \$1 when the data is perfect, the cost of it with defective data would be \$10. The total cost for 100 records, as discussed in the example above, can be calculated as:

$$\text{Total Cost} = (62 \times \$1) + (38 \times \$10) = \$62 + \$380 = \$442$$

This calculation shows how defective data records cost about four times more, compared to if the data was perfect.

- * **The Rule of Ten** is supported by multiple studies from Japan, the USA, and Great Britain, which dealt with the causes of product and quality defects. Even though the studies focused on manufacturing processes, the consequences can be found in modern software development as well. If it takes €100 to fix a defect at unit testing, it takes €1,000 at system testing, €10,000 at Acceptance Testing, and €100,000 after release.