* 1. **Data Quality Reference Model**

**1. What Is a Data Quality Reference Model?**

A **Data Quality Reference Model** provides a structured way to understand, measure, and manage the **dimensions of data quality** within an organization. It serves as a **framework** to assess how well data meets business needs and supports decision-making.

According to DAMA-DMBOK V2, the reference model includes:

1. **Data Quality Dimensions**
2. **Associated Business Rules or Requirements**
3. **Measurement Techniques**
4. **Remediation Methods**
5. **Ongoing Monitoring and Governance**

**2. Core Components of the DQ Reference Model**

| **Component** | **Explanation** |
| --- | --- |
| **Dimension** | Key characteristic of data (e.g., accuracy, completeness, consistency) |
| **Definition** | Description of what that dimension means in business context |
| **Rules/Criteria** | Specific rules to evaluate each dimension (e.g., no null values) |
| **Metrics** | Quantitative measures used to monitor data quality (e.g., error rate) |
| **Target Thresholds** | Acceptable levels for each metric (e.g., 98% accuracy) |
| **Remediation Approach** | Methods to correct or prevent data quality issues |

**3. Example: Dimensions in the DQ Reference Model**

| **Dimension** | **Definition** | **Example Rule** |
| --- | --- | --- |
| Accuracy | Data correctly represents real-world values | Address must match postal registry |
| Completeness | Required data is present | No nulls in mandatory fields |
| Consistency | Data is the same across systems | Account type should match in CRM and billing |
| Uniqueness | No duplicate records | Customer ID should be unique |
| Timeliness | Data is available when needed | Orders updated within 15 minutes of creation |
| Validity | Data conforms to rules and formats | Email must match regex pattern |
| Integrity | Relationships are correctly maintained | Foreign key values must exist in parent table |

**4. How to Apply the Data Quality Reference Model**

**Step 1: Identify Critical Data Elements (CDEs)**

Focus on data elements that impact key business outcomes.

* Example: Customer name, date of birth, email, national ID

**Step 2: Define Quality Rules per Dimension**

Define what “good data” looks like for each CDE.

* Example:
  + **Completeness**: Email must not be null
  + **Accuracy**: DOB must match external ID registry
  + **Uniqueness**: National ID must not be duplicated

**Step 3: Profile the Data**

Use tools to analyze current data against defined rules.

* Example tools: Talend, Informatica, Python scripts

**Step 4: Measure and Score**

Assign scores (e.g., 95% accurate, 90% complete) and compare to thresholds.

**Step 5: Report Findings**

Visualize quality issues with dashboards (Power BI, Tableau)

**Step 6: Remediate**

* Fix invalid data (e.g., use trusted source to correct DOB)
* Update processes to prevent future errors (e.g., mandatory email field at input)

**Step 7: Monitor Continuously**

Schedule monthly checks and track trends over time

**5. Real-World Scenario: Applying the DQ Reference Model**

**Organization: MedSure Health Insurance**

**Goal: Improve data quality of member records to reduce claim rejection rates**

**Step-by-Step Application**

**Step 1: Identify Critical Data Elements**

* Member ID
* Policy Type
* Date of Birth
* Hospital Name

**Step 2: Define Rules per Dimension**

| **Element** | **Dimension** | **Rule** |
| --- | --- | --- |
| Member ID | Uniqueness | Must be unique across the system |
| Date of Birth | Accuracy | Must match the government ID system |
| Policy Type | Validity | Must be one of {Silver, Gold, Platinum} |
| Hospital Name | Completeness | Must not be null on any claim |

**Step 3: Profile Data**

Results:

* 2.5% of Member IDs duplicated
* 8% of DOBs mismatched with national ID
* 5% of claims missing hospital names

**Step 4: Score and Report**

* Accuracy score: 92%
* Completeness: 95%
* Uniqueness: 97%

**Step 5: Remediate**

* Cleaned duplicate IDs
* Integrated ID validation API at registration
* Made hospital name a required field in the claims form

**Step 6: Monitor**

* Setup weekly data quality jobs
* Data steward reviews exception reports

**Outcome:**

* Claim rejection rate dropped by 38%
* Customer satisfaction improved due to fewer delays
* Audit compliance scores increased