* 1. **The different facets of data quality, and why validity is often confused with quality**

**Facets of Data Quality – As per DAMA-DMBOK V2**

DAMA-DMBOK V2 outlines several **key dimensions (facets)** of data quality that must be assessed to determine if data is **fit for purpose**. These dimensions reflect different characteristics of data that impact its usefulness.

**1. Accuracy**

**Definition**: The degree to which data correctly describes the real-world object or event it represents.

**Example**:  
A customer’s date of birth is recorded as “1990-05-17” but their actual birthdate is “1990-07-17”.  
This is inaccurate data, even though the format is valid.

**2. Completeness**

**Definition**: The extent to which required data is present and not missing.

**Example**:  
An insurance claim record is missing the claimant’s contact number.  
The record is incomplete, and this may delay claim processing.

**3. Consistency**

**Definition**: The degree to which data is the same across different datasets or systems.

**Example**:  
A customer’s address is “123 Main St” in the CRM but “321 Main St” in the billing system.  
This inconsistency may lead to mail delivery issues or customer frustration.

**4. Validity**

**Definition**: The extent to which data conforms to the correct format, structure, or domain constraints.

**Example**:  
An email address “john.doe@example.com” is valid because it matches the expected format.  
But “john.doe@example” would be invalid due to a missing domain suffix.

**5. Uniqueness**

**Definition**: The degree to which each record or data element is distinct and not duplicated.

**Example**:  
Two identical customer records exist with the same ID and contact info.  
This duplication violates uniqueness and may cause confusion in reporting.

**6. Timeliness**

**Definition**: The extent to which data is up-to-date and available when needed.

**Example**:  
A sales report is using last month’s data instead of today’s.  
The report is timely only if updated as per business expectations.

**7. Integrity**

**Definition**: The extent to which data relationships are correctly maintained.

**Example**:  
A foreign key in an order table references a non-existent customer ID.  
This breaks referential integrity and causes logic errors in queries.

**Why Validity Is Often Confused with Overall Data Quality**

Many users—especially those with a technical or data entry focus—assume that **if data is valid**, it must be **high quality**. However, **validity is only one part of data quality**, and **valid data can still be wrong, incomplete, or misleading**.

**Detailed Examples to Clarify the Difference**

**Example 1: Valid but Not Accurate**

* **Email field**: “john.doe@gmail.com” → This is **valid format**
* But if John actually uses “john.doe@company.com”, the value is **not accurate**

**Lesson**: A field can pass validation checks but still misrepresent the truth.

**Example 2: Valid but Not Complete**

* **Form field**: Phone number = “+1-202-555-0176” → Format is **valid**
* **However**, the email field is blank

**Lesson**: One valid field does not mean the **record as a whole** is complete or usable.

**Example 3: Valid but Not Unique**

* Two customer records:
  + Name: Alice Smith
  + Email: alice.smith@example.com
* Both are **valid and formatted correctly**, but they represent **the same person**.

**Lesson**: Valid entries that are duplicated may cause confusion and reporting errors.

**Example 4: Valid but Not Consistent**

* System A: Product category = “Electronics”
* System B: Product category = “ELECTRONICS” or “Elec”

All values are **valid**, but inconsistent in naming, which causes problems in integration and analytics.

**Key Takeaways**

* **Validity** is about **format and domain rules**, not meaning or usefulness.
* Validity checks are **necessary but not sufficient** to ensure high-quality data.
* **True data quality** requires **multiple facets to work together**: accuracy, completeness, uniqueness, consistency, timeliness, and integrity.

**Best Practices to Avoid Confusion**

* Train users to understand the **difference between format checks and quality assessments**
* Use **data profiling tools** that check for **accuracy, completeness, and duplication** alongside validation
* Implement **data quality scorecards** that cover **all dimensions**, not just validity
* Include **data stewards** in processes to ensure that data meets **business-level expectations**