* 1. **Metadata Expanded Explanation**

**“Metadata describes the data itself, the concepts the data represents, and the connections between the data and concepts.”**

This captures the **three core roles** of metadata in data management:

**1. Metadata Describes the Data Itself**

This includes **technical details** about the structure and storage of the data.

**Examples**:

* Table name: customer\_info
* Column: customer\_id (INTEGER, PRIMARY KEY)
* Data type: String, Date, Number
* Source system: CRM Database
* File format: CSV, JSON, XML

**Purpose**:  
Helps systems and engineers understand how to store, query, and process the data.

**2. Metadata Describes the Concepts the Data Represents**

This includes **business context** and **semantic meaning**, which makes data understandable to non-technical users.

**Examples**:

* customer\_id: "A unique identifier assigned to each customer"
* net\_revenue: "Total sales minus discounts and returns"
* Business owner: Finance department
* Definition source: Business glossary

**Purpose**:  
Helps analysts, stewards, and decision-makers interpret the data correctly.

**3. Metadata Shows Connections Between Data and Concepts**

This is where **relationships and data lineage** come into play—metadata links data elements to business terms, systems, processes, and other data.

**Examples**:

* Lineage: “net\_revenue” in dashboard → comes from “revenue\_table” → derived from “sales\_data” and “returns\_table”
* Relationships: “customer\_id” in orders table links to “customer\_id” in customer\_info table
* Governance link: net\_revenue metric is governed by Finance, validated weekly

**Purpose**:  
Enables:

* Impact analysis
* Traceability (for audits)
* Regulatory compliance
* Cross-system integration

**Summary**

| **Aspect** | **Metadata Description** |
| --- | --- |
| **Data itself** | Technical metadata (format, type, storage) |
| **Concepts the data represents** | Business metadata (definitions, meanings, policies) |
| **Connections and relationships** | Lineage, data flow, relationships between entities |

**Real-World Example**

**Term**: Customer Lifetime Value (CLV)

* **Data Description**: clv\_score, FLOAT, stored in customer\_metrics table
* **Concept Description**: “Projected total revenue a customer will generate during their relationship with the company”
* **Connections**: Derived from purchase\_history, returns, and loyalty\_program data; owned by Marketing; refreshed weekly

**Metadata Scorecard Calculations:**

**1. Metadata Completeness**

**KPI**: Percentage of data assets with complete metadata.

**Formula**:  
(Number of assets with required metadata / Total number of assets) × 100

**Example**:  
If out of 1,000 data assets, 850 have all required metadata (e.g., data owner, description, data type),  
then completeness = (850 / 1000) × 100 = **85%**

**2. Metadata Accuracy**

**KPI**: Percentage of metadata entries that are correct and up to date.

**Formula**:  
(Number of accurate metadata entries / Total metadata entries) × 100

**Example**:  
If 900 of 1,000 metadata records have correct and current information,  
accuracy = (900 / 1000) × 100 = **90%**

**3. Metadata Timeliness**

**KPI**: Percentage of metadata records updated within the defined time frame (e.g., last 6 months).

**Formula**:  
(Number of timely metadata updates / Total metadata entries) × 100

**Example**:  
If only 600 of 1,000 records were updated in the last 6 months,  
timeliness = (600 / 1000) × 100 = **60%**

**4. Metadata Usage**

**KPI**: Number of times metadata is accessed or queried by users (e.g., in a catalog or search).

**Formula**:  
Count of user queries or metadata views per month.

**Example**:  
If users accessed metadata 1,200 times in a month,  
usage KPI = **1,200 views/month**

**5. Metadata Coverage**



**KPI**: Percentage of critical data elements (CDEs) documented with metadata.

**Formula**:  
(Documented CDEs / Total CDEs identified) × 100



**Example**:  
If 400 of 500 critical data elements have metadata,  
coverage = (400 / 500) × 100 = **80%**



**6. Data Stewardship Assignment**

**KPI**: Percentage of data assets with an assigned data steward or owner.

**Formula**:  
(Assets with assigned owner / Total assets) × 100

**Example**:  
If 750 of 1,000 assets have owners,  
ownership KPI = (750 / 1000) × 100 = **75%**

**Metadata Scorecard Example**

| **KPI** | **Target** | **Actual** | **Status** |
| --- | --- | --- | --- |
| Metadata Completeness | 90% | 85% | ⚠️ At Risk |
| Metadata Accuracy | 95% | 90% | ⚠️ At Risk |
| Metadata Timeliness | 80% | 60% | ❌ Poor |
| Metadata Usage | 1000/mo | 1200 | ✅ Good |
| Metadata Coverage (CDEs) | 100% | 80% | ⚠️ At Risk |
| Steward Assignment | 100% | 75% | ⚠️ At Risk |