**Master and Reference Data Management**

**Master Data**

* **Definition:**  
  Master data represents the **core business entities** that are critical to operations. These are the key “things” an organization cares about and uses repeatedly across different systems and processes.
* **Examples:**
  + Customers
  + Products
  + Employees
  + Suppliers
  + Accounts
* **Characteristics:**
  + Usually **stable** but can change over time (e.g., a customer’s address).
  + Shared and used by multiple business processes and systems.
  + Requires strong governance for consistency and quality.

**Master Data**

**Definition:**  
Master data represents the **core business entities** that are relatively stable over time and are used repeatedly across transactions and processes.

**Examples:**



| **Entity Type** | **Master Data Example** |
| --- | --- |
| **Customer** | Customer ID, name, address, contact details, loyalty tier |
| **Product** | Product SKU, name, description, category, price |
| **Employee** | Employee ID, name, department, role |
| **Supplier** | Supplier ID, company name, contact info |
| **Location** | Store ID, city, region, warehouse location |

**Characteristics:**

* Relatively static (not changed frequently)
* Shared across multiple systems (e.g., CRM, ERP, sales)
* Used to provide context to transactions

**Normal (Transactional or Operational) Data**

**Definition:**  
Normal data (often called **transactional data**) represents **day-to-day operations and activities**. It is dynamic and typically references master data.

**Examples:**

| **Transaction Type** | **Normal Data Example** |
| --- | --- |
| **Sales Order** | Order ID, date, quantity, total price, **Customer ID**, **Product ID** |
| **Invoice** | Invoice number, date, amount, **Customer ID** |
| **Shipment Record** | Shipment ID, date, shipping status, **Product SKU**, **Warehouse ID** |
| **Employee Timesheet** | Timesheet ID, date, hours worked, **Employee ID** |

**Characteristics:**

* Volatile (changes frequently)
* Time-stamped
* Depends on master data for context
* Drives operational processes

**Master Data vs Normal Data – In a Sales Scenario**

| **Aspect** | **Master Data** | **Normal Data** |
| --- | --- | --- |
| **Customer** | John Smith, Customer ID: C123 | Order placed by C123 on May 12 for $250 |
| **Product** | Product SKU: P456, “Wireless Mouse” | 3 units of P456 sold in Order O789 |
| **Employee** | Jane Doe, Sales Rep, ID: E321 | E321 logged 8 hours on May 12 |

**Why It Matters**

* **Master Data** ensures consistent definitions (e.g., what a "product" is).
* **Normal Data** captures what actually happens in day-to-day business (e.g., what was sold, when, and to whom).

**Reference Data**

* **Definition:**  
  Reference data is a **set of predefined, static values** used to **classify or categorize** other data. It acts as a standard list of permissible values.
* **Examples:**
  + Country codes (e.g., US, IN, FR)
  + Currency codes (e.g., USD, EUR, INR)
  + Industry codes
  + Gender codes (e.g., M, F, O)
  + Status codes (e.g., Active, Inactive)
* **Characteristics:**
  + Typically **small**, **static**, and **standardized**.
  + Changes infrequently.
  + Used to ensure data consistency and validation.

**Summary Table**

| **Aspect** | **Master Data** | **Reference Data** |
| --- | --- | --- |
| Represents | Core business entities | Standardized classifications |
| Examples | Customers, Products, Employees | Country codes, Currency codes |
| Change Frequency | Moderate (changes over time) | Rarely changes |
| Purpose | Used for operational transactions | Used for categorization & validation |
| Governance | High importance for accuracy & consistency | Usually controlled centrally |

**Master Data Scenario: Customer and Product**

Imagine a retail company managing sales:

* **Master Data Example:**
  + **Customer Master Data:** Each customer has a unique ID, name, address, contact details, and loyalty status.
  + **Product Master Data:** Each product has a unique SKU, name, description, price, and category.
* **Usage:**  
  When a customer places an order, the system references the **customer master data** to validate customer details and the **product master data** to identify the items sold.
* **Why Important:**  
  Consistent and accurate master data ensures the company knows who its customers are and what products are being sold, avoiding mistakes like shipping errors or billing to the wrong customer.

**Reference Data Scenario: Country Codes and Order Status**

In the same retail company:

* **Reference Data Example:**
  + **Country Codes:** Customers select their country from a predefined list (e.g., US, CA, UK) when entering shipping details.
  + **Order Status Codes:** Orders can be "Pending", "Shipped", "Delivered", or "Cancelled".
* **Usage:**
  + Country codes standardize addresses, enabling correct shipping and reporting.
  + Order statuses categorize the progress of each order in a consistent way across all systems.
* **Why Important:**  
  Using reference data prevents data entry errors and makes reporting and analytics reliable because everyone uses the same standard values.

**Summary:**

| **Data Type** | **Business Scenario** | **Impact** |
| --- | --- | --- |
| Master Data | Customer and Product records | Accurate transactions, customer service, inventory control |
| Reference Data | Country codes for shipping, order status codes | Consistent data input, accurate reporting, operational efficiency |

**2. Master Data Management (MDM)**

**2.1 Definition**

Master Data Management is the process of creating and maintaining a single, accurate, and consistent source of core business entities. These entities are used across multiple systems and business units.

**2.2 Objectives of MDM**

* Ensure consistency and accuracy of critical data
* Eliminate duplicate and conflicting data
* Provide a trusted source of data for enterprise use
* Improve reporting and decision-making

**2.3 Examples of Master Data**

* Customer information (name, contact details)
* Product details (name, category, pricing)
* Employee records (ID, role, department)
* Supplier data (name, location, contact)

**2.4 Key MDM Activities**

* Identification of core master data entities
* Establishing authoritative data sources
* Data profiling, cleaning, and standardization
* Matching and merging duplicate records
* Applying business rules and survivorship logic
* Distributing mastered data across systems
* Assigning data stewardship and governance

**2.5 MDM Use Case Example**

An organization has customer data in the sales, support, and billing systems. MDM is used to match and merge these records into a single customer profile that can be trusted and reused across the company.

**3. Reference Data Management (RDM)**

**3.1 Definition**

Reference Data Management is the process of managing controlled lists or standardized values that classify or categorize other data. It supports consistency and regulatory compliance.

**3.2 Objectives of RDM**

* Maintain standard and approved value sets
* Ensure data categorization is consistent
* Support interoperability across systems
* Align with industry and external standards

**3.3 Examples of Reference Data**

* Country codes (e.g., US, UK, IN)
* Currency codes (e.g., USD, EUR, INR)
* Product status (e.g., Active, Discontinued)
* Payment methods (e.g., Credit Card, Bank Transfer)

**3.4 Key RDM Activities**

* Define and document code lists and hierarchies
* Validate and approve reference values
* Align with external standards (e.g., ISO)
* Manage versions and historical changes
* Distribute reference data to consuming systems

**3.5 RDM Use Case Example**

A company maintains a list of countries and currency codes that are used across billing, logistics, and CRM systems. Reference Data Management ensures that all departments use the same, correct values.

**4. Key Roles in MDM and RDM**

* Data Stewards: Manage day-to-day data quality and standards
* Data Owners: Define data requirements and approve changes
* Data Architects: Design data models and integration methods
* IT Administrators: Support MDM and RDM tools and infrastructure

**5. Common Challenges**

* Lack of data ownership and accountability
* Duplicate or inconsistent data across systems
* Difficulty integrating data from multiple sources
* Resistance to centralized data governance
* Inadequate tools for managing data

**6. Success Factors**

* Strong executive sponsorship
* Clear business use cases and goals
* Defined data governance and stewardship
* Continuous data quality monitoring
* Use of appropriate tools and technologies

**7. Technologies and Tools**

* MDM platforms: Informatica MDM, SAP Master Data Governance, Oracle MDM
* RDM tools: Collibra, Ataccama, Talend, IBM InfoSphere

**8. Summary Table**

|  |  |  |
| --- | --- | --- |
| **Area** | **Master Data Management** | **Reference Data Management** |
| Purpose | Manage core business entities | Manage classification code sets |
| Data Examples | Customers, Products, Employees | Country codes, Currency types |
| Activities | Match, merge, sync, govern | Define, validate, distribute |
| Goal | Trusted, consistent business data | Standardized values across systems |
| Tool Support | MDM hubs/platforms | Reference data catalogs or repositories |

**9. Conclusion**

Master and Reference Data Management are critical to achieving data consistency, quality, and governance across the organization. By applying the principles from DAMA DMBOK v2, organizations can ensure that their shared data assets are well-defined, trustworthy, and support strategic business outcomes.