

ARTICLE 4

Data Governance and Master Data Management

Nenad Jukić
Loyola University Chicago

In this article, we will give a brief overview of data governance and master data management.

Data Governance

Data governance is a broad term describing a set of initiatives within an organization aimed at formally regulating how, when, and by whom the organizational data and metadata are created, stored, updated (inserted, modified, deleted), and archived.

Even though data is one of most important assets (if not the most important) in many organizations, it is often not treated with the same level of organizational rigor as other assets, such as real estate, money, or vehicles, which are more traditionally perceived and appreciated as resources. However, more and more corporations are becoming aware of the importance of a regulated approach toward the data. This awareness usually results in implementing data governance.

The goal of data governance is to establish and enforce proper rules and policies for the treatment of data. By doing so, organizations formally regulate the management of data similar to the way they regulate the management of their other important assets, such as financial resources. For example, most corporations have very firm rules about who is allowed to use the company's money, for what purposes, on which occasions, by whose permission, and so on. Data governance institutes similar rules for dealing with corporate data.

An example of data governance is the development and implementation of the database policies and standards described in Chapter 11 of *Database Systems: Introduction to Databases and Data Warehouses (Edition 3.0)*. Another example of a data governance initiative is creating policies that require the undertaking of the preventive and corrective data quality actions described in Chapter 6 of *Database Systems: Introduction to Databases and Data Warehouses (Edition 3.0)*. In many cases, data governance initiatives are related to the broader laws and regulations that govern the actions of the organization. For example, US health care organizations have to align their data management practices with the Health Insurance Portability and Accountability Act (HIPAA) imposed by the US federal government. Among other provisions, this act contains strict privacy rules that regulate the use and disclosure of patients' information. In US health care organizations, these rules must be implemented in database systems that host patient information as part of their overall data governance framework.

In most organizations the role of a **data steward** encompasses tasks related to data governance. Data stewards are responsible for the proper use of the data in databases. Typical data stewardship activities include the following:

- the creation and enforcement of policies or standards for appropriate data entry, update, and use
- the creation and implementation of data quality control activities
- the creation, implementation, and maintenance of business metadata (such as business rules and descriptions of tables and columns)
- other activities ensuring compliance with data governance policies

In addition to the role of data steward, many organizations also establish the role of data custodian. A **data custodian** is responsible for the technical aspects of data management and use, such as the protection, transport, and storage of data. In many organizations, there is a significant overlap between the roles of a data steward, a data custodian, and/or a database administrator (DBA). The distribution of the titles and associated responsibilities vary broadly from organization to organization.

Master Data Management

Master data management (MDM) is one of the most common organizational data governance initiatives. **Master data** in an organization contains an authenticated quality version of the key data that provides a common point of reference for the organization's information systems. Once the master data is in place, operational information systems within the organization align their data with the master data.

For example, the ZAGI Retail Company could maintain the table MASTERPRODUCT, containing information about all products that are sold by the ZAGI Retail Company. All information systems within the ZAGI Retail Company that use product information would have to verify that their product information is consistent with the product information in the MASTERPRODUCT table.

Master data management involves creating and maintaining a collection of quality master data tables and ensuring that the use of master data is embedded across the operational information systems in the organization. It is essential that the master data is of the highest quality. Master data must exhibit all the properties of high-quality data (accuracy, uniqueness, completeness, consistency, timeliness, and conformity) described in Chapter 6 of *Database Systems: Introduction to Databases and Data Warehouses (Edition 3.0)*.

Not all data in operational information systems is master data. However, the data that corresponds to master data must be aligned with the master data. Consider the example where a store clerk in the ZAGI Retail Company working in a store records in the sales management information system that a certain customer bought three items of a certain product at 8:00 a.m. The quantity (three) and time (8:00 a.m.) are not related to the master data. However, assuming the existence of the master tables MASTERPRODUCT, MASTERCUSTOMER, and MASTERSTORE, the data about *product*, *customer*, and *store* in the sales information system have to be verified against and aligned with the data in these master tables.

The main benefit of proper master data is that the key information in use throughout the organization is consistent. The concept of master data for operational systems is similar to the concept of conformed dimensions for data warehouses discussed in Chapter 8 of *Database Systems: Introduction to Databases and Data Warehouses (Edition 3.0)*. Whereas conformed dimensions provide a set of quality reference data for consistent analytical use, master data provide a set of quality reference data for consistent operational use.

There are various architectures used for the management of master data. Three main approaches are the following:

- centralized approach
- registry
- hybrid approach

In a centralized approach, a single, central copy of the master data is used by operational information systems. Nonmaster data is still collected and maintained by the operational information systems, but all instances of master data are retrieved from and updated in the central master data copy. This centralized approach is illustrated in Figure 1.

An opposite approach to centralized master data is keeping all master data in various individual operational information systems. Such dispersed master data is connected via a central **master data registry**. A master data registry contains only a list of keys that are used to connect and coordinate the actual master data that resides in operational systems. The master data registry allows any individual information system to access master data from other information systems in order to align and supplement its own master data. A master data registry is illustrated by Figure 2.

The hybrid approach to storing master data combines the two approaches described previously. There is an actual central master data copy, but individual operational systems may contain their own copies of the master data as well. The central master data copy and other master data copies are connected and coordinated via the central master data registry. This hybrid approach is illustrated in Figure 3.

In some implementations of the hybrid approach, master data can be updated (inserted, deleted, or modified) only in the central copy and all the updates are propagated to the copies in the operational systems. Other implementations allow master data to be updated in individual operational systems. These changes are propagated to the central copy and then to the other individual systems.

Also, in some implementations, the entire set of master data is present in the central copy and in all individual systems, while in other implementations, portions of the master data are present in the individual systems and/or in the central copy.

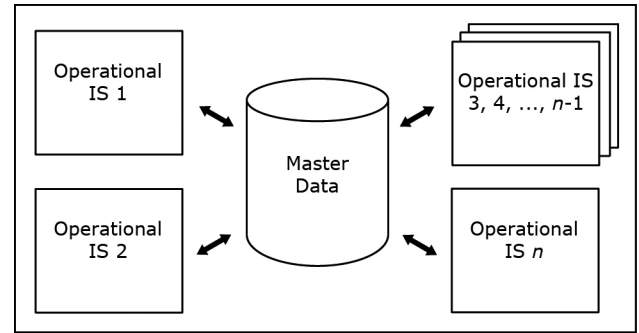


Figure 1 Centralized master data.

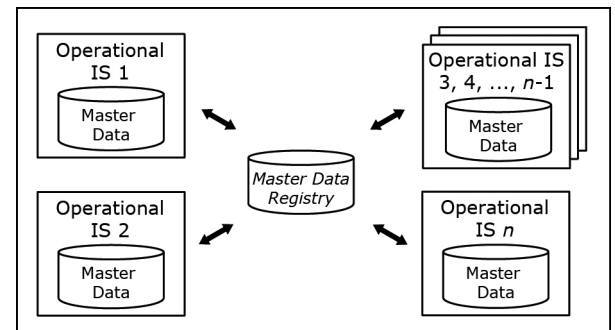


Figure 2 Master data registry.

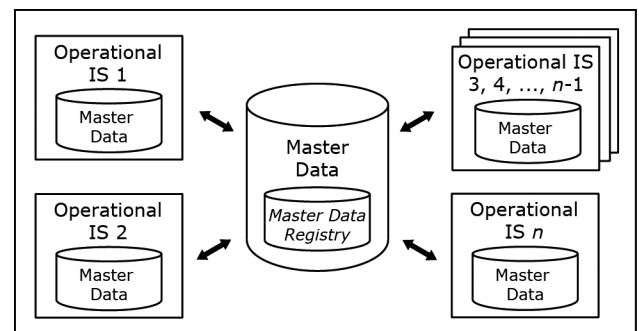


Figure 3 A hybrid MDM approach.