A Quick Guide to Data Dictionary

What is a Data Dictionary?

A data dictionary[1] is a collection of descriptions of data objects or terms, definitions, and properties in a data asset. Data dictionaries provide information about the data.

Data dictionaries are an essential communications tool for data modeling, curation, governance, and analytics, especially when dealing with datasets that have been collected, compiled, categorized, used, and reused by different internal and external data Consumers across the organization.

Why do you need a data dictionary?

Data dictionaries provide valuable definitions for the data as well as help data Consumers understand any data asset before delving into the details within that asset. In order for data to be trusted and appropriately used, it must be understood and supported by clear definitions.

An established data dictionary can provide organizations many benefits, including:

- Greater data standardization and consistency across the organization or domain
- Improved analysis and decision-making based on better understanding of data
- Improved documentation
- Increased reuse of data

What is in a Data Dictionary?

A data dictionary consists of several data components, which contains multiple levels: **data asset**, **entity**, **attribute**, and **value domain**. Each level includes different components, but each component should be defined with the following properties:

Data Component Name	Data Component Name that represents a class of real-world entities or characteristics of those entities
Description	A short description for the data component name
Type	Logical or physical data component
Required	Required or optional data component
Sample	A sample of the data component

The data asset level is composed of one line that contains a data asset profile, which includes the data asset name, description, type, version, and create and last update date.

[1] Refer to the CMS Data Description Guidelines webpage to download the Data Dictionary Template.

The entity level is composed of one line per entity/table that contains all required properties for this element, such as logical and physical data element name, requirement ID (justification for having this entity in the system), security category (Consumers that can access this entity) and so forth.

The attribute level is composed of one line per attribute/column that includes logical and physical data element name (as well as parent element, i.e. entity/table name) and some logical and physical properties (data type and size, null option, is attribute/column a primary key, a foreign key, an unique/alternative key, PII field, etc.).

Finally, the value domain level is composed of a list of groups/collections of similar values (sex codes, account types, credit card types, enrollment plans, etc.) defined as business rules, look-up tables or list of valid values, and each available value within each collection.

Logical vs. Physical Data Dictionary

A logical data dictionary describes information in business terms and focuses on the meaning of terms and their relationship with other terms. A physical data dictionary represents data in a specific database and includes actual tables and columns in the data asset's schema.

- **Logical:** Business-related names, entities/attributes, logical naming convention (mixed case). Platform-agnostic.
- **Physical:** Technical names, tables/fields, platform-driven naming convention, field length and data types. Platform-specific.

How do you create and maintain a Data Dictionary?

Most data modeling tools and database management systems (DBMS) have built-in, active data dictionaries the capable of generating and maintaining data dictionaries. Data stewards may also utilize the CMS Data Dictionary Template[2] and guide for manually creating a simple data dictionary in Excel.

Best Practices:

- Start building a data dictionary during the gathering business requirements phase.
- The data dictionary is a living document that must be regularly maintained.
- If utilizing erwin, then Consumers should build the data dictionary from their data model using Report Designer.
- Reference the CMS Data Naming Quick Reference Guide[3] for data naming guidance.
- Reference the CMS Data Definition Quick Reference Guide[4] for data definition guidance.

[2] Refer to the CMS Data Description Guidelines webpage to download the Data Dictionary Template.
[3] For additional data naming guidance, refer to the CMS Data Naming Quick Reference Guide.
[4] For additional data definition guidance, refer to the CMS Data Definition Quick Reference Guide.