

A Quick Guide to Data Dictionary

What is a Data Dictionary?

A data dictionary is a collection of descriptions of data objects or terms, definitions and properties in a data model or database. Data dictionaries provide information about your data.

Data dictionaries provides valuable definitions for the data and help users understand any dataset before diving into it. That makes them 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 people across the organization.

In the *Data Naming Guidelines* document, the ‘Data Component Name’ can be used for an object term, a property term and if necessary, a representation term. Please see *Data Naming Guidelines* document for further guidance.

Why do you need a Data Dictionary?

For data to be trusted, it needs to be understood. It needs to have a definition that everyone agrees upon. An established data dictionary can provide organizations many benefits, including:

- Improved data quality
- Improved trust in data integrity
- Improved documentation and control
- Reduced data redundancy
- Reuse of data
- Greater data standardization and consistency across the organization
- Easier data analysis
- Improved decision making based on better understanding of data
- Enforcement of standards

What’s in a Data Dictionary?

To create a simple data dictionary, below is the minimum set of data elements you should look to include.

Data Component Name	The name of the data component that represents a class of real-world entities or concepts.
Description	A short description for the data component name.
Data Type	A data type or value set of a data component. E.g., CHAR, DATE, INTEGER, DECIMAL etc.

A data dictionary may include other data elements to suit the needs of the organization. Additional data elements may include data format, maximum length, where the data was collected from (the source), whether it holds sensitive data, who owns the data, any caveats of its use, or any other data that is deemed important to have in the dictionary.

Logical vs. Physical Data Dictionary

Logical data dictionary describes information in business terms and focuses on the meaning of terms and their relationship with other terms while a physical data dictionary represent data in a specific database. They include actual tables and columns in your database schema.

¹ See *CMS Data Naming Guidelines*, CMS Data Architecture and Engineering Services, 2020. [Online]. Available: <https://www.cms.gov/ActualURLToBeUpdatedWhenPublished>

A Quick Guide to Data Dictionary

- Logical: Business-related names (compatible with Business Glossary), 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 and have the capability to generate and maintain data dictionaries. You may also use CMS data dictionary template and guide for manually creating a simple data dictionary in Excel. [**Note:** See *CMS Data Dictionary Template*]

Best Practices:

- Start building a data dictionary right when you are gathering business requirements.
- The data dictionary is a living document. Regularly maintain your data dictionary.