Data Dictionary

IBM Dictionary of computing defines data dictionary as

"A centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format"

A data dictionary is also referred to as metadata repository. It comprises of information about the primitive and composite data elements.

A landline number is an example of a composite data element and it comprises of primitive data elements like ISD Code, STD Code, Area code etc.

Let's consider a complete telephone number in India, it looks as follows:

+91 22 2857 1200

This number comprises of multiple information:

Basic Element Name	Any other name	Type of values	Description
ISD Code	Country Code	Number	The international dialling code for the country
STD Code	City Code	Number	The country's city code
Area Code	Area Code	Number	Local telephone exchange number
Number	Telephone Number	Number	Actual phone number
Telephone Number =			ISD Code + STD Code + Area Code + Number

The table above describes primitive data elements and the composite data elements. This table is an example of data dictionary.

Purpose of Data Dictionary

Data dictionary is used to create a standardized representation of data elements for a given project or projects and for its stakeholders. This helps in having a common understanding and representation.

Without having a data dictionary, data can have multiple interpretations in a system. This may lead to different representations. For example, one developer may present a phone

number as one value (allowing numbers and '+' sign) and other one may present it as a combination of multiple fields. This leads to bad user experience.

Please note that data dictionary is always created for a context (a project). A composite data element may be relevant for one project but may not be for the other.

For example, if the application under development is for customers across the world, the ISD code is a relevant part of the telephone number. We have created the data dictionary for an application with ISD number.

But, if the application is to be used only in one country, the ISD code has no relevance and should not be part of the data dictionary.

Formats for Data Dictionary

Multiple formats can be used to represent a data dictionary. One of the formats is used on the previous page. BABOK v3 guide uses the following format:

Primitive Data Elements	Data Element 1	Data Element 2	Data Element 3
Name Name referenced by data elements	First Name	Middle Name	Last Name
Alias Alternate name referenced by stakeholders	Given Name	Middle Name	Surname
Values/Meanings Enumerated list or description of data element	Minimum 2 characters	Can be omitted	Minimum 2 characters
Description Definition	First Name	Middle Name	Family Name
Composite	**Customer Name** = First Name + Middle Name + Family Name		

Elements of Data Dictionary

As per BABOK v3 guide, the primitive data elements should have the following information:

- **Name:** A meaningful name to the element so that it be recognized uniquely.
- **Aliases:** Alternate or commonly used term used for the data element, different from Name
- **Value/Meanings:** Represents data types (like number or text), size, list of values (if applicable) etc

- **Description:** More details about the data elements for the specific project/context

Composite Elements may be:

- **Sequences:** A composite element comprises of multiple primitive data elements. These elements could be arranged in a sequence i.e. coming one after the other as in the telephone number. A '+' sign is used to represent that as shown below:

  [ISD Code + STD Code + Area Code + Telephone Number]

- **Repetitions:** When one of the data elements gets repeated multiple times. A '{ }' is used to represent that:

  Invoice = Invoice_number + Date + {items-ordered} + Sub-total + Sales-Tax + Total due

- **Optional Elements:** A composite data element may also have optional elements. Optional elements are represented by '( )'.

  Customer_name = First_name + (Middle_name) + Last_name

Note: Care must be taken to create the data dictionaries as unnecessary and irrelevant data dictionaries can lead to limited or no realization of business value.