

2.3.4 Data Taxonomy

[return to Taxonomic Views](#)

Data, although on the surface appears to be a simple, straight forward concept that “everyone” understands, it is actually a complex concept with multiple facets and interpretations. Data is generally considered a plural term with the singular form being **datum**, however, common usage allows for data to be considered as singular with the singular-plural differentiation being restricted to academic and scientific usage. ¹⁾

A more formal definition of data is provided by Organisation for Economic Co-operation and Development (OECD)²⁾:

Characteristics or information, usually numerical, that are collected through observation.

In essence, Data are individual elemental, facts, statistics, or items of information, often numeric, that are collected through observation. Technically, data are a set of values of qualitative or quantitative values about one or more persons or objects,³⁾ while a datum (singular of data) is a single value of a single variable.⁴⁾

Unfortunately, the terms “data” and “information” are in common use interchangeable. However, these terms have distinct meanings and are part of a larger framework defining an increasing ordered set of cognitive layers. (See: [xapend.i_cog_model](#)).

In some popular publications, data are sometimes said to be transformed into information when they are viewed in context or in post-analysis.[3] However, in academic treatments of the subject data are simply units of information. Data are used in scientific research, businesses management (e.g., sales data, revenue, profits, stock price), finance, governance (e.g., crime rates, unemployment rates, literacy rates), and in virtually every other form of human organizational activity (e.g., censuses of the number of homeless people by non-profit organizations).

Data are measured, collected, reported, and analyzed, and used to create data visualizations such as graphs, tables or images. Data as a general concept refers to the fact that some existing information or knowledge is represented or coded in some form suitable for better usage or processing. Raw data (“unprocessed data”) is a collection of numbers or characters before it has been “cleaned” and corrected by researchers. Raw data needs to be corrected to remove outliers or obvious instrument or data entry errors (e.g., a thermometer reading from an outdoor Arctic location recording a tropical temperature). Data processing commonly occurs by stages, and the “processed data” from one stage may be considered the “raw data” of the next stage. Field data is raw data that is collected in an uncontrolled “in situ” environment. Experimental data is data that is generated within the context of a scientific investigation by observation and recording.

Data has been described as the new oil of the digital economy.[4][5]

There are several taxonomies that cover Data within a DIDO. The taxonomies are not additive in nature rather than mutually exclusive. For example, any data within the DIDO can have a Cognitive classification

as well as a state and or a digital taxonomy. A piece of data can be classified as **Knowledge, At Rest**, and a **Coin**; it might just as easily been classified as **Understanding, In Motion**, a **Non-Fungible asset**.

The three taxonomies are described in detailed in the following sections:

- [2.3.4.1 Cognitive Taxonomy of Data](#)
- [2.3.4.2 State of Data Taxonomy](#)
- [2.3.4.3 Type of Data Taxonomy](#)

Data is a complex concept and unfortunately the term is overloaded. Generically, it can describe all a range of concepts from a simple data point to the complex range of data and its association. In the [Cognitive Model](#) data is just one level in the human cognitive hierachy.

In any data project, there are likely to be two components. The first is the data collected, assembled, or generated. Think of it as the raw content in the system. It could be hourly temperature readings from a sensor, the age of individuals in a survey, recordings of individual voices, or photographs of plant specimens. The second component is the data system in which the data is stored and managed.⁵⁾

We usually do not think of data content separate from the system in which it is stored, but the distinction is important in terms of intellectual property rights. The question is what, if anything, is protected by [Copyright](#). Data that is factual has no copyright protection under U.S. law; it is not possible to copyright facts. Not all data is in the public domain. A project might, for example, use copyrighted photographs; the photographs are part of the project's "data." In many cases, the data in a data management system as well as the metadata describing that data will be factual, and hence not protected by copyright.

A database, on the other hand, can have a thin layer of copyright protection. Deciding what data needs to be included in a database, how to organize the data, and how to relate different data elements are all creative decisions that may receive copyright protection.

Because of the different copyright status of databases and data content, different mechanisms are required to manage each. Copyright can govern the use of databases and some data content (that which is itself original), but contract law, trademarks, and other mechanisms are required to regulate factual data.

The third word in the acronym DIDO is Data and every [node](#) therefore, manages and controls data. The data within the node is classified as either: [ledger data](#), [ancillary data](#) or [external data](#).

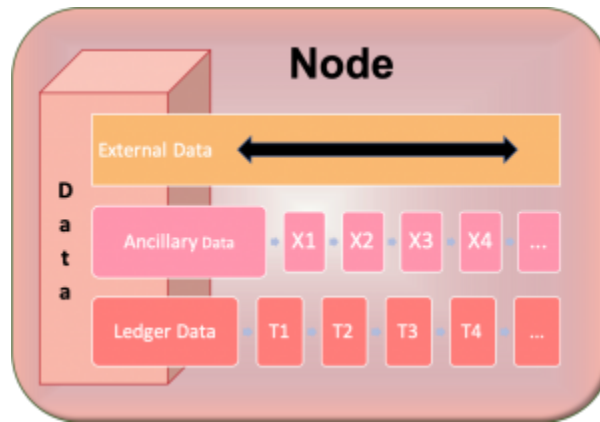


Figure 1: Types of Node Data

- 2.3.4.1 Cognitive Taxonomy of Data
- 2.3.4.2 State of Data Taxonomy
- 2.3.4.3 Type of Data Taxonomy
- 2.3.4.4 Data Model Taxonomy
- 2.3.4.5 Data Lifecycle Taxonomy
- 2.3.4.6 Geographic Jurisdiction Data Governance Taxonomy
- 2.3.4.7 Data Memory Taxonomy
- 2.3.4.8 Object Data Taxonomy
- 2.3.4.9 Error Taxonomy

1)
Data Is or Are: Is Data Singular or Plural?, The Editor's Manual, Accessed: 28 September 2021,
<https://editorsmanual.com/articles/data-singular-or-plural/>

2) , 3)
DATA, Organisation for Economic Co-operation and Development (OECD), Accessed: 28 September 2021,
<https://stats.oecd.org/glossary/detail.asp?ID=532>

4)
Data, Wikipedia, Accessed: 28 September 2021, https://en.wikipedia.org/wiki/Data#cite_note-:0-1

5)
Introduction to intellectual property rights in data management, Cornell University, Research Data Management Service Group, Accessed: 26 September 2021,
<https://data.research.cornell.edu/content/intellectual-property>

From:
<https://www.omgwiki.org/dido/> - **DIDO Wiki**

Permanent link:
https://www.omgwiki.org/dido/doku.php?id=dido:public:ra:1.2_views:3_taxonomic:4_data_tax:start&rev=1632856203

Last update: **2021/09/28 15:10**

