

Writing Enterprise-Quality Data Definitions

Tips for Creating Terms and Definitions



by Gwen Thomas
The Data Governance Institute

The Data Governance Institute
www.DataGovernance.com
2511 E. Colonial Drive #219
Orlando, FL 32803 USA
telephone: 321.438.0774



Writing Enterprise-Quality Data Definitions

What happens if silo-quality data definitions are used in an enterprise project?
Typically, one of three scenarios:

- ❑ You get lucky.
 - Assumptions are made by a cross-functional group of analysts, architects, system developers, DBAs, testers, and project stakeholders. Luckily, they all make the same assumption, and it turns out to be right. The data definition that was used by one silo within the enterprise is adequate to the needs of all, and it doesn't contradict any other meanings.
- ❑ You get unlucky.
 - A data element actually means one thing, but some project members assume that it means something else. This has created an issue, but it remains hidden.
 - The issue is uncovered during testing.
 - Now the project team has to spend time understanding what's going on. After understanding the issue, they have to find a resolution. They need to conduct an impact analysis to determine what other problems with the system were caused by this issue. Sometimes they need to tear apart the system to fix problems, or they have to introduce patches that could compromise the system later. They need to develop new test strategies, plans, and scripts.
 - Eventually the project makes it back to where it was before someone had to "stop the assembly line" to diagnose and fix the problem. Other projects whose assembly lines were stopped while team members worked this problem can get started again.
- ❑ You get really unlucky.
 - A data element actually means one thing, but some project members assume that it means something else. This has created an issue, but it remains hidden.
 - Assumptions are made, and the issue finds its way into an application, process, service, or system. For some reason it isn't caught during testing.
 - The process, service, or system is deployed, only to deliver bad data.
 - Business decisions are made based on bad data.



The work of defining data

We may think of writing data definitions as a task assigned to teams working with Data Architecture/Modeling, Metadata Management, or the within a System Development Lifecycle. However, this same effort may be undertaken by other groups working with business terms, such as

- Business Process Modeling
- Taxonomy Development
- Web Personalization or Search Engine Optimization
- Information Architecture (structure, headings, menu items, navigation links, headings, or hyperlinks) for web sites, documents, or presentations
- Report Generation / Data Analysis / Business Intelligence
- Business Rules Management

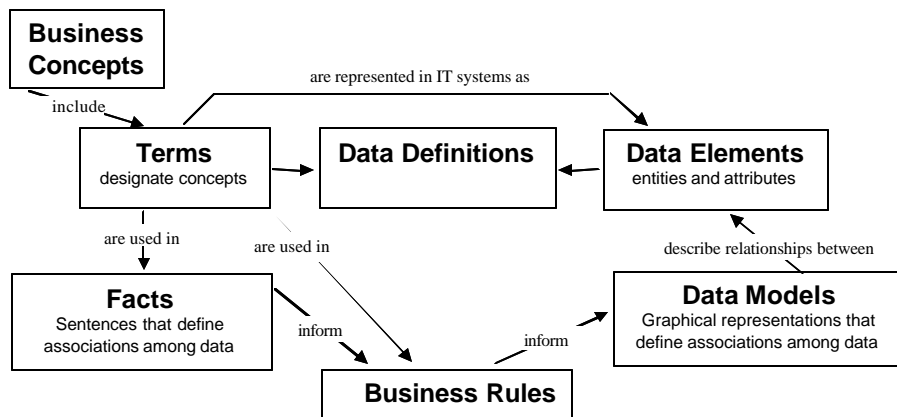
In this case, the work may be labeled as defining business terms or business rules. Still, this is data work, as many business terms will eventually be represented in the organization's data architecture. Also, the nouns in business rule statements must be traceable to data, and that data must be unambiguously defined.

So, whether we're focusing on business terms or data elements, we must create clear, concise, and useful definitions.

Terms Used in This Paper

- A **rule** is a declarative statement that applies logic or computation to information values. It results in either a decision about taking action or the discovery of new information.
- A **business rule** is a set of conditions, determined by business leaders, that govern the manner in which business is conducted.
- In the context of running a business – or documenting business rules – a **term** is a noun or noun phrase that designates a business concept. Often, a term is represented an organization's data architecture by a **data element**, or **field**.
- A **data model** is a graphical representation of the relationships between data elements.
- In the context of business rules management, a **fact** is a business concept that is held to be true, expressed as one or more sentences that define associations among data.

Comparing Business Rules/Terms and Data Modeling/Definitions



Enterprise-Quality Data Definitions

Enterprise-quality data definitions are adequate for use in cross-functional processes, services, and controls. They bring clarity and consistency to system integration projects, data warehouses, service-oriented architectures, and other enterprise projects. Enterprise-quality terms and their definitions are:

- Atomic
- Traceable
- Consistent
- Accurate
- Clear
- Complete
- Concise
- Declarative

Atomic terms

An atomic term describes one concept only, generally a person, place, thing, idea, or event.

- While a term can have only one definition, that definition can be shared by more than one term, if they are synonyms. If synonyms are used in your environment, your dictionary or repository should contain those synonyms, clearly indicating that they point to the same data element.
- While a term describes a single concept, it may take more than one word to name that concept. A term can be:
 - A simple concept that can be named with a single word.
Example: *Address*.
 - A simple concept that can only be named with more than one word.
Example: *Street Address*
 - A slightly more complex concept that consists of a simpler term preceded by a modifier. Examples: *Delivery Address*, *Primary Address*, *Previous Address*.
In this case, the simpler term should be defined, also, if any of the following apply:
 - A meaning could be assumed by some within the business. In this case, it's important to define it to avoid incorrect assumptions.
 - It is not used by the business precisely because of its ambiguity or history of confusing usage.
Example: *Cartoon* is a legacy name that has been replaced by multiple current terms, such as *Animated Feature*, *Animated Short*, and *Comic Strip*.

Traceable terms

A traceable term can be mapped to a particular data element.

- A business term can map to
 - *An entity (Example: Customer)*
 - *An attribute of an entity (Example: Customer Name)*
 - *A value in a domain or reference data set (Example: Female)*
 - *A set of values or reference data (Example: Customer Type)*



- ❑ Often, the same name has been assigned to data elements in multiple systems across the enterprise. Ideally, these data elements should represent the same business concept, and therefore can share the same definition. If, however, discrepancies in meaning exist, then the definition of the data element should indicate the true and precise business term being represented by the data element.

Consistent definitions

Consistent definitions employ a standardized syntax or structure, as well as consistent conventions to point out context.

- ❑ If the term is used within a certain context, that context is specifically called out. (Example: In the context of Lean Manufacturing or the Six Sigma quality methodology, Waste in a process is any activity that does not result in moving the process closer to the final output or adding value to the final output.)
- ❑ Likewise, all terms within the collection which have special meaning because of a context use the same convention to point out that context. (Example: Within the SAP system, ...)

This requirement is especially important when you are moving from silo definitions to enterprise definitions. A context that might have been obvious to users within a business or technical silo might easily be lost to wider audiences.)

Accurate definitions

An accurate definition is precise and correct for every situation in which a reader of the definition would employ the term. Disclaimers or usage notes are included as needed. When multiple definitions for a word exists, the one that applies to the business term or data definition is clearly designated.

- ❑ Again, watch out for silo definitions. These are definitions that are well accepted within the context of one group or application but might have different meanings to different groups or in different applications.
- ❑ Typically, don't try to resolve silo definitions by making one group use the other group's definition. Instead, create and define a new term that is a synonym. Add the new term and also the original silo terms to the repository, with the silo terms clearly labeled as legacy terms.

Clear definitions

A clear definition should be understandable by the reader. It should avoid using highly-technical or obscure language.

- ❑ Be as precise as possible.
- ❑ Avoid circular definitions where terms define each other. Also, avoid "cheeseburger" definitions (A cheeseburger is defined as a burger with cheese.)
- ❑ If possible, avoid possessive language in the definition of the term. Instead, spell out relationships.



Complete definitions

A complete definition contains the term, a verb such as MEANS or IS, and a description. If needed, it also includes language to put the term in context.

Concise definitions

A concise definition does not contain extraneous information, adjectives, or flowery language.

Declarative definitions

When a definition exists outside of a dictionary format (such as this list), it is expressed as a declarative sentence.

- ❑ The subject of the sentence is the name of the term. The definition does not include an implied subject.
- ❑ The subject begins the sentence, unless a clause is needed to establish context. Example: CUST_CURR_M_ADDR1 is the first line of the customer's current mailing address. The definition is not written in passive tense (The first line of the customer's current mailing address is expressed as CUST_CURR_M_ADDR1).

Summary

Data definitions that are adequate for use in a single silo within an organization may not meet the needs of the entire enterprise. You can help ensure that your data definitions are enterprise-quality by ensuring that they are Atomic, Traceable, Consistent, Accurate, Clear, Complete, Concise, and Declarative.



About the Data Governance Institute

The Data Governance Institute is the premier provider of in-depth, vendor-neutral information about – and assistance with – tools, techniques, models, and best practices for the governance of data and information.

The Institute provides a wealth of resources: the free DGI Data Governance Framework, information on data laws, regulations, and standards, whitepapers, case studies, best practices, data humor, and non-technical briefings on data-related issues and disciplines.

The Institute also provides community for governance practitioners, and it directly assists clients through training, issue analysis, and program assistance as they roadmap, design, and implement governance programs.

