

Data Definitions: Guidelines and Examples

“You cannot effectively exchange what you cannot understand.”

This document summarizes guidelines for defining data components with semantic clarity and simplicity.

Defining data components by applying guidelines consistently makes it easier for users who aren't familiar with said components to understand, use or re-use, and benefit from them.

1. Define a data component, not its name.

A data definition is a representation of a concept, via a descriptive phrase, which distinguishes the concept from related concepts.

Define a data component before naming it. Reversing this sequence may lead to a definition that precisely describes the name but does not adequately describe the concept the data component represents.

2. Unless the underlying concept itself is plural, define a data component in the singular.

In the following incorrect example, it's unclear whether a reference number applies to one article or several. Avoid confusion by consistently defining a data component in terms of a single instance of said component.

Incorrect Example: Article Number: Reference number identifying articles.

Correct Example: Article Number: A reference number that identifies an article.

3. Define a data component in terms of what it is, not only what it is not.

In the following incorrect example, the “negative” definition leaves unclear what the data component actually represents.

Incorrect Example: Freight Cost Amount: Costs which are not related to packaging, documentation, loading, unloading, and insurance.

Correct Example: Freight Cost Amount: Cost amount incurred by a shipper in moving goods from one place to another.

4. Define a data component in a descriptive phrase or sentence(s).

A clearly-written explanation is almost always necessary to precisely define a concept and avoid ambiguity.

Incorrect Example: Agent Name: Representative.

Correct Example: Agent Name: A name of a party authorized to act on behalf of another party.

5. Expand uncommon abbreviations on their first occurrence.

Many abbreviations are not commonly known outside of specific contexts. Use the full term of an abbreviation to enhance understanding.

Incorrect Example: Tide Height: A vertical distance from MSL to a specific tide level.

Correct Example: Tide Height: A vertical distance from mean sea level (MSL) to a specific tide level.

6. Define a data component concisely, with only the level of detail needed to state the essential meaning of the underlying concept.

In the following incorrect example, the inclusion of extraneous material renders the definition less clear than the following correct example. The additional verbiage does not enhance understanding.

Incorrect Example: Invoice Amount: A total sum of all chargeable items mentioned on an invoice, taking into account deductions on one hand, such as allowances and discounts, and additions on the other, such as charges for insurance, transport, handling, etc.

Correct Example: Invoice Amount: A total sum charged on an invoice.

In the following incorrect example, the additional language does not enhance understanding of the underlying concept in the present context.

Incorrect Example: Character Set Name: A name given to the set of phonetic or ideographic symbols in which data is encoded, for the purpose of this metadata registry, or, as used elsewhere, the capability of systems hardware and software to process data encoded in one or more scripts.

Correct Example: Character Set Name: A name for a set of phonetic or ideographic symbols in which data is encoded.

7. A data component's definition should be precise, unambiguous, and allow only one possible interpretation.

In the following incorrect example, it is unclear what is meant by "delivered." A definition should make explicit what the underlying concept is.

Incorrect Example: Shipment Receipt Date: A date on which a specific shipment is delivered.

Correct Example: Shipment Receipt Date: A date on which a shipment is received by the receiving party.

8. A data component's definition should stand alone.

In the following incorrect example, the definition unnecessarily requires the aid of a second definition to explain the meaning of the first.

Incorrect Example: School Location City Name: See "School Site".

Correct Example: School Location City Name: A name of a city where a school is situated.

9. Define a data component without embedding other definitions.

Definitions should only describe the data component at hand. If a term within a definition requires its own definition, it should be defined separately.

Incorrect Example: Sample Type Code: A code identifying the kind of sample collected. A sample is a small specimen taken for testing.

Correct Example: Sample Type Code: A code identifying the kind of sample.

10. Define a data component without using circular reasoning.

In the following examples, two data components are defined by referring to each other. The reader does not gain an understanding of either of the underlying concepts.

Circular Example, Pt 1: Employee ID Number: A number assigned to an **employee**.

Circular Example, Pt 2: Employee: A person corresponding to an **employee ID number**.

11. Use consistent phrasing and logical structure for related definitions.

Using common phrasing for similar or associated concepts enhances the association(s) for readers.

Consistent Example, Pt 1: GoodsDispatchDate: A date on which goods were **dispatched** by a given party.

Consistent Example, Pt 2: GoodsReceiptDate: A date on which goods were **received** by a given party.

12. Define a data component without embedding rationale, functional usage, or procedural information.

A good definition just explains the underlying concept a data component represents. Additional information, while useful in many cases, should be provided in other metadata attributes.

Incorrect Example: Data Field Label: Identification of a field in an index, thesaurus, query, database, etc., which is provided for units of information such as abstracts within tables.

Correct Example: Data Field Label: An identification of a field in an index, thesaurus, query, database, etc.

13. Define a data component without re-defining its constituent terms.

When defining a data component with multiple terms, i.e., when defining a property of an object or entity, avoid describing the data component in a way that's inconsistent with the definition of the object or entity.

It's okay to use a name's terms within a corresponding definition to enhance clarity.

Example (object): Person: A human being.

Incorrect Example (property): Person Hair Color: A color of the hair of a data modeler.

Correct Example (property): Person Hair Color: A color of the hair of a person.