WIKIPEDIA

Data element definition

In metadata, a **data element definition** is a human readable phrase or sentence associated with a data element within a data dictionary that describes the meaning or semantics of a data element.

Data element definitions are critical for external users of any data system. Good definitions can dramatically ease the process of mapping one set of data into another set of data. This is a core feature of distributed computing and intelligent agent development.

There are several guidelines that should be followed when creating high-quality data element definitions.

Contents

Properties of clear definitions

A good definition is:

1. **Precise** - The definition should use words that have a precise meaning. Try to avoid words that have multiple meanings or multiple word senses.
2. **Concise** - The definition should use the shortest description possible that is still clear.
3. **Non Circular** - The definition should not use the term you are trying to define in the definition itself. This is known as a circular definition.
4. **Distinct** - The definition should differentiate a data element from other data elements. This process is called disambiguation.
5. **Unencumbered** - The definition should be free of embedded rationale, functional usage, domain information, or procedural information.

A data element definition is a required property when adding data elements to a metadata registry.

Definitions should not refer to terms or concepts that might be misinterpreted by others or that have different meanings based on the context of a situation. Definitions should not contain acronyms that are not clearly defined or linked to other precise definitions.

If one is creating a large number of data elements, all the definitions should be consistent with related concepts.

Critical Data Element – Not all data elements are of equal importance or value to an organization. A key metadata property of an element is categorizing the data as a Critical Data Element (**CDE**). This categorization provides focus for data governance and data quality. An organization often has various sub-categories of CDEs, based on use of the data. e.g.,

1. Security Coverage – data elements that are categorized as personal health information or PHI warrant particular attention for security and access
2. Marketing Department Usage – the marketing department could have a particular set of CDEs identified for identifying Unique Customer or for Campaign Management
3. Finance Department Usage – the Finance department could have a different set of CDEs from Marketing. They are focused on data elements which provide measures and metrics for fiscal reporting

Standards such as the ISO/IEC 11179 Metadata Registry specification give guidelines for creating precise data element definitions. Specifically chapter four of the ISO/IEC 11179 metadata registry standard covers data element definition quality standards [1] (http://standards.iso.org/ittf/PubliclyAvailableStandards/c035346_ISO_IEC_11179-4_2004(E).zip).

Using precise words

Common words such as **play** or **run** frequently have many meanings. For example the WordNet (http://wordnet.princeton.edu/) database documents over 57 different distinct meanings for the word "play" but only a single definition for the term **dramatic play**. Fewer definitions in a chosen word's dictionary entry is preferable. This minimizes misinterpretation related to a reader's context and background. The process of finding a good meaning of a word is called Word sense disambiguation.

Examples of definitions that could be improved

Here is the definition of "person" data element as defined in the www.w3c.org Friend of a Friend specification * (http://xmlns.com/foaf/0.1/#term_Person):

```
Person: A person.
```

Although most people do have an intuitive understanding of what a person is, the definition has much room for improvement. The first problem is that the definition is circular. Note that this definition really does not help most readers and needs to be clarified.

Here is the definition of the "Person" Data Element in the Global Justice XML Data Model 3.0 * (http://gjxdmtools.gtri.gatech.edu/ssgt/subset?propertyKey=km-115):

```
person: Describes inherent and frequently associated characteristics of a person.
```

Note that once again the definition is still circular. Person should not reference itself. The definition should use terms other than person to describe what a person is.

Here is a more precise but shorter definition of a person:

```
Person: An individual human being.
```

Note that it uses the word *individual* to state that this is an instance of a class of things called human being. Technically you might use "homo sapiens" in your definition, but more people are familiar with the term "human being" than "homo sapiens," so commonly used terms, if they are still precise, are always preferred.

Sometimes your system may have cultural norms and assumptions in the definitions. For example if your "Person" data element tracked characters in a science fiction series that included aliens you may need a more general term other than *human being*.

```
Person: An individual of a sentient species.
```

See also

- Data dictionary
- Data element
- Global Justice XML Data Model
- NIEM
- ISO/IEC 11179
- Metadata
- Metadata registry

References

Sources

1. ISO/IEC 11179-4:2004 Metadata registries (MDR) - Part 4 (http://standards.iso.org/ittf/PubliclyAvailableStandards/c03 5346_ISO_IEC_11179-4_2004(E).zip)
2. ISO/IEC Technical Report 20943-1, First edition, 2003-08-01 Information technology — Procedures for achieving metadata registry consistency

Retrieved from "https://en.wikipedia.org/w/index.php?title=Data_element_definition&oldid=774929295"

This page was last edited on 11 April 2017, at 15:32.