



Business Data Vocabulary Analysis Process

1. INTRODUCTION

At the start of the database development for a systems automation project, the business will provide a business requirements document (BRD). The business requirements are the description of the company's accounting practices, that is, how the company implemented the axiomatic system of double entry bookkeeping with journals, registers and ledgers. Specifically, it identifies the business process to be automated and names the business data objects that are involved in the process.

The business data object names used depend on

- company history – the names will be the names in general use when the company was founded. For example, in the landline telephone industry, historically, one phone number is one account is one customer. To bill one customer for many phone numbers in a single bill requires special processing.
- company business model – the completeness of the company business model will depend on whether a single business data object has more than one data concept. (Generalization Mistakes)
- company culture – how the company sees itself will determine whether or not the data concepts can be generalized to meet new requirements.

The business data vocabulary analysis process executes the following tasks:

1. Document the business data terms.
2. Define each business data term
3. Capture the relationships between the business data terms

The objective of documenting the business data vocabulary is to enable the business to share its understandings of its data concepts and to document how the business names its data. There will be many names for a business data concept depending on where in the business process the conversation is taking place. But the databases used for recordkeeping will use only a single name for this business data concept.

The business vocabulary documentation must be able to resolve three main business issues:

1. Synonyms – business data objects that have different names but the same meanings
2. Homonyms – business data objects. that have the same names but different meanings
3. Improve Data Quality to reduce the need for:
 - Data Cleansing – checking data values when data is moved
 - Data Redundancy – keeping derived data values
 - Data Duplication – keeping multiple copies of data structures and values
 - Data Reconciliation – explaining why similar reports contain different data values

The goal of maintaining a business data vocabulary document is to develop the inventory of data objects and the processes that use those data objects as sources and sinks of data for the Information Technology organization. The business vocabulary documentation enables the IT organization to meet the business requirement for describing the data, reporting its use and tracking the transformations as the data moves from source to destination. (Data Life Cycle)

2. INPUT

At the start of the database development project, the business will provide a BRD. The BRD contains the description of what is required in the automated system, the business process names, and the business data terms that the business uses to name the data.

Also, the project will have a list of business stakeholders, contact information and their roles in the process being automated.

The assumptions made at the start of the project are that the BRD is comprehensive, complete and unambiguous.

3. BUSINESS DATA TERMS

The first step is to parse BRD to identify the business data terms named (list of business data terms). The business data terms should be nouns in the BRD. Technically, this process is known as extracting the Universe of Discourse (Simsion).

The business data terms are collections of data used in the business processes. The individual data items within the collections of data are identified and documented in the logical data modeling step of the DDLC.

For each item in the list of business data terms, document the

- Name – the name or label that the business uses for the data object
- Data term type – the type of data object. One of input form (paper or data capture screen), output from (printed report or reporting/inquiry screen), etc.
- Source name – the name of the file in which the data object was found
- Source location – the location of the document that contains the data object definition.
- Description – additional comments about the data object.

Each business stakeholder should validate that the list has no missing business data terms and the documentation for each term is complete and accurate.

4. DEFINITIONS

The next step in documenting the business vocabulary is to formalize the definition of the data objects. For each item in the list of business data terms, get the definition of the item by describing:

- Business meaning – the definition or meaning of the business data term
- Data life cycle phase – the place in the data life cycle where the business data term occurs. One of data capture, data reconciliation, or data presentation.
- Processing phase – the place in the business process where the business data term is used. One of product initiation, supply chain, manufacturing, sales, fulfillment, or get payment.
- Definition source – the name of the business unit that owns the business data object
- Data steward – the name of the person responsible for the business data object
- Document location – the physical location of the documentation for this business data object
- Description – additional comments about the data object.

Each business stakeholder should provide a definition for each item in the list of business data terms.

If you find that a business data term has two disparate definitions, i.e., the term means one thing to the first business unit and it means something different to a second business unit, then you have found a homonym in the business vocabulary. Homonyms will cause problems in the implementation of the database where every label is unique and has a single meaning. The business stakeholders must decide how they want to handle homonyms.

5. SYNONYMS

The next step in documenting the business vocabulary is to document the business thesaurus. This is done in two steps. The first step is to document all the synonyms of the list of business data terms.

Each business stakeholder should provide a list of synonyms for each item in the list of business data terms. Each synonym must be an element of the list of business data terms.

For example, the business term worker has the synonyms of employee, associate, contractor, etc. The synonyms will have exactly the same definition and scope.

6. SUBTYPES

The second step is to document all the subtypes of the list of business data terms.

Each business stakeholder should provide a list of subtypes for each item in the list of business data terms. Each subtype must be an element of the list of business data terms.

For example, in the health care industry, the business term employee has the subtypes of doctor, nurse, receptionist, clerk, etc. The subtypes will have different definitions and less scope than that of the supertype.

7. DELIVERABLES

At this point, the business data vocabulary analysis is complete.

The deliverables required to close out the analysis process are:

1. Business Vocabulary Dictionary
 - a. List of business data terms
 - b. List of business data term definitions
2. Business Vocabulary Thesaurus
 - a. List of business data term synonyms
 - b. List of business data term subtypes

8. REFERENCES

Generalization Mistakes – <http://thedataorg.com/Page03/Generalization%20-%20A%20modeling%20problem.pdf>

Data Life Cycle – <http://thedataorg.com/Page09/DataLifeCycleArchitecture.pdf>

Simsion, Graeme C., Data Modeling Theory and Practice, Technics Publications, LLC, Bradley Beach, New Jersey, 2007