

Data Item Naming Standard

Background

What is a data item? It can be defined as the most basic logical unit of data. A naming standard for data items is important. Why? Because a standard:

- Promotes consistency of use across systems.
- Helps us to identify, document, and manage our information assets.
- Improves the accuracy of searches for a particular piece of data.
- Can improve communication by representing the organization's vocabulary of terms.

Our old data item (for the purposes of this document the term data item refers to columns of tables and attributes of entities) naming standard was based on a standard developed years ago by the National Institute of Standards and Technology (NIST). NIST developed a standard, based on the standards developed at AT&T and many other companies. A name started with a prime word, then had one or more modifiers and ended with the class word. It also standardized the base class word list. The standard eliminated the variety of punctuation marks and insisted on underscores, hyphens, or blanks.

Our standard modified this only slightly and was essentially prime word-modifier-modifier-class word. Class word was only required for readability:

- A prime word describes the object being identified
- A modifier (or qualifier) describes the data element itself
- A class word defines the type or purpose of an attribute (e.g., date, amount)

The Problem

Over time, problems emerged with this standard:

- The prime word is redundant in data models.
- Technology imposed constraints on format (e.g. name length, allowed characters, and case sensitivity) for database columns. Constraints change over time.
- Attitudes and sensibilities change over time.

General Principles

- A name should be unique (when fully qualified)
- A name should be meaningful
- A name should describe what the data item is
- A name should be independent of the application itself and independent of hardware and software used, i.e., the name should be independent of physical implementation characteristics

Naming Standard Proposal

- Use the whole name if feasible and reasonable.
- Length constraints of the target environment may be limiting.
- Have a standard list of abbreviations to be used where the whole name is not used. The abbreviation should uniquely identify its associated term, but still be easy enough to remember when encountered in a name. Data Administration will manage the list of abbreviations. Abbreviations will be from two to five characters in length.

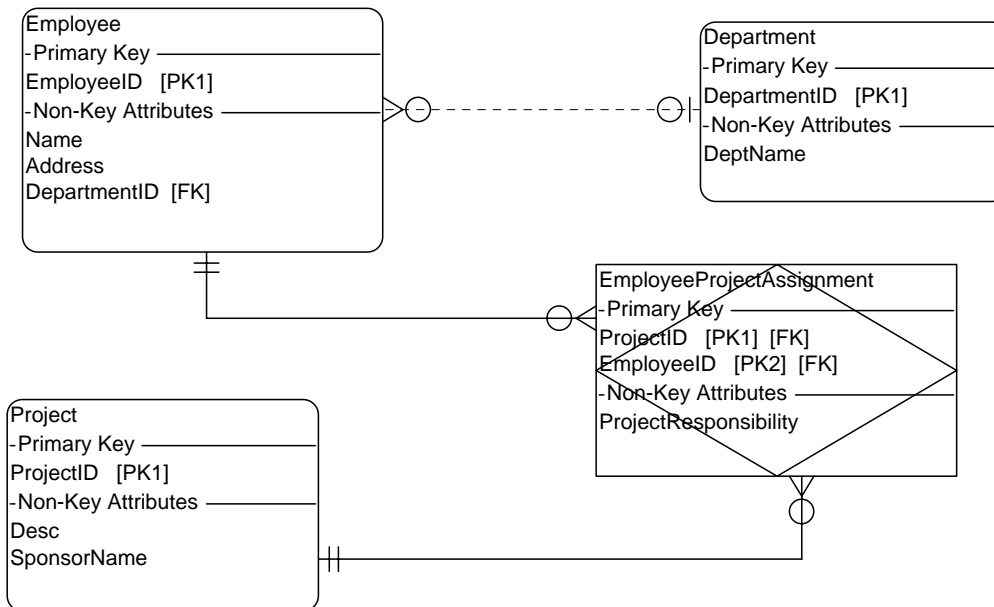
Data Item Naming Standard

- Acronyms should be used in all cases where applicable. An acronym constitutes a standard abbreviation. SIC (Standard Industrial Classification) is an example. It is possible to have acronyms with more than one meaning. This is acceptable because it should be relatively rare and obvious from the context as to the meaning intended.
- Ordinarily, do not include table or entity names in the data element names. Exceptions to this rule: primary key (unique identifier) and foreign key fields; where clarity is needed; or, by common usage (see Example).
- Capitalize the first letter of each word in the name; acronyms should normally be capitalized.
- Attribute name (logical) should be the same as column name (physical).
- Aliases should not be used except for "business names" of data items.
- A "business name" for a data item is whatever the organizational unit calls a data item. This is a reflection of reality and the richness of the English language. The business name or names should be recorded and linked to the data item in the repository. One department may call a governmental unit allowed to levy taxes a tax authority and another call it a tax district. The data item should be named consistently, but with references to both business names. The challenge for us is recognizing that these are different business names for the same object, which should be described in the repository only once.
- Have a "core" set of organizational data elements with standard names, properties, and components. Examples of these would be name, address, and phone number.
- These logical data item names would map directly to physical model (schema) names; however, if denormalization is done and tables are combined, then names may need to be changed to avoid naming conflicts.

[Example is on next page]

Data Item Naming Standard

Example. Logical data model:



In the example we can see:

- "ID" is appended to the name of the entity for the unique identifier of each entity. ID is the standard abbreviation or acronym for identifier. Department + ID = DepartmentID.
- Name is the standard for a "person" type entity's name.
- DeptName is used to differentiate a department's name from an individual's. It has different properties, such as no last name, no first name.
- Address is the standard for geographic location data.
- Desc is used as a standard abbreviation for a descriptive text about an entity.
- ProjectResponsibility is a fully spelled-out name.
- Foreign keys match the name of the primary key entity, and we can see DepartmentID in the Employee entity matching DepartmentID in the Department entity.