SYBASE®
An SAP Company

WHITE PAPER

Building a Better Data Dictionary with PowerDesigner®

HOW MODELING ADDS VALUE TO A DATA DICTIONARY

As organizations move to managing information as a first class corporate asset, the primary challenge will not be in just finding the right sources of the information from the right systems, but in identifying and attributing the right meaning to that data. Data dictionaries allow organizations to standardize data content, context, and definitions to achieve consistency and reusability as well as drive quality information into new or existing tactical and strategic business initiatives.

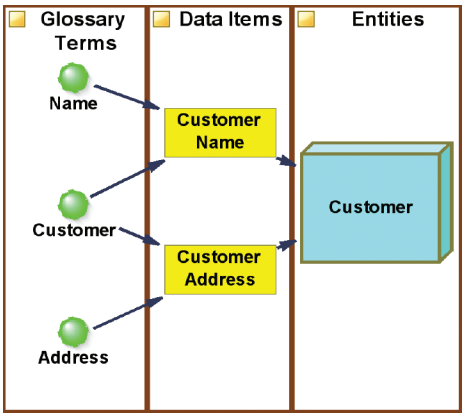


Figure 1 - Terms to Data Items to Entities

A data dictionary will provide:

- easier integration and communication between systems
- more standardized messaging between applications
- higher quality business intelligence and analytics
- better understanding between all subject matter experts

The building blocks of the data dictionary include business glossary terms, their use to define names for data elements, data items and their use as entity attributes, entity relationships and the definitions for each. Data dictionaries ensure consistency of use throughout the organization by providing a common business language, and "single version of the truth" for all common data elements used across the enterprise.

Best practices to consider when establishing a data dictionary:

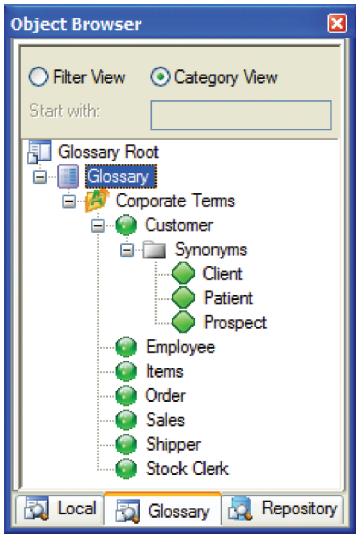
- As the content of the data dictionary mature over time, version control becomes important. Keeping track of what version of the data dictionary a project has derived its data implementation from will help others to understand why definitions may differ and how to rectify differences during integration, federation and business intelligence transformation projects. Repositories will be very helpful in managing versions and configurations of models, to ensure that old versions remain referenced.
- Data dictionaries will transform over time, and a good design will accommodate that. Best practices include not overloading any single concept, making it easy to reuse and easy to maintain.
- If a concept becomes obsolete, do not reuse it to avoid confusion with legacy implementations. It is important to allow for any data concept to expand over time, and therefore important to have a clear alignment between data entities and their attributes. Normalization processes are used to achieve this alignment.
- Taxonomies and related terminologies will differ between business units, organizations and companies. Minimize mappings across all these different terminologies. It is nearly impossible to avoid having a need to share a common definition to different dictionary terms. Use aliases or synonyms to support the need to have different names for the same data concept. and mappings between data concepts.

REQUIREMENTS FOR THE DATA DICTIONARY

Preparing the data dictionary needs to take certain key criteria into account. The following list identifies the most common requirements for supporting the storage and maintenance of the data dictionary, but additionally includes the workflow and processes that support the creation and management of the information.

Elements of a Data Dictionary:

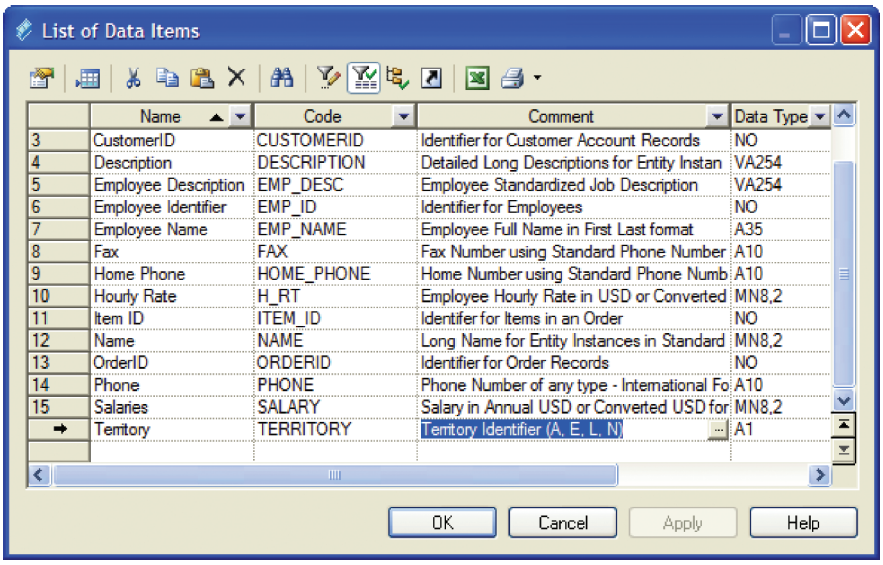
- A hierarchical list of business glossary terms, aliases and relationships between terms
- A unique list of entities and data items
- Descriptions of data artifacts
- Entity-attribute relationships resulting from a data item being assigned to describe an entity or entities
- Entity to entity relationships



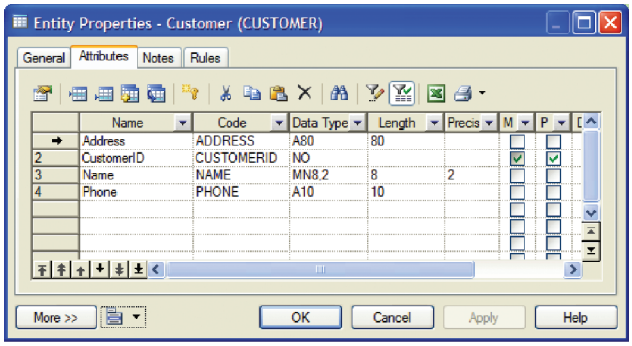
A hierarchical list of business glossary terms: To improve business alignment, PowerDesigner's Enterprise Glossary ensures business terms and definitions are established and consistently managed throughout all models to ensure everyone is speaking 'the same language'. Glossary terms are defined within a hierarchical classification system, and terms may relate to other terms. PowerDesigner's Enterprise Glossary offers "preferred" terms for data identification whenever a new or alias is encountered to keep the number of instantiations of terms down to a minimum.

A unique list of entities and data items: PowerDesigner's metamodel maintains a single definition of each dictionary object (ex: there is only one definition of 'customer') no matter where it is subsequently used throughout the enterprise. PowerDesigner® allows for a single concept (model, diagram) to be defined once, and reused wherever needed as appropriate.

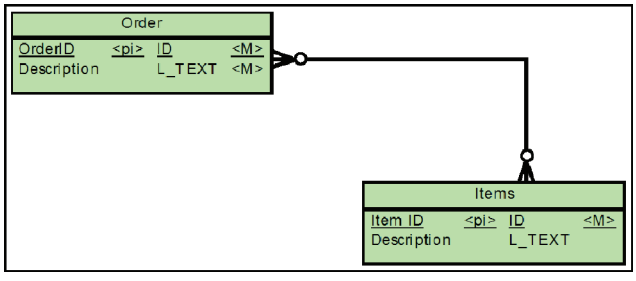
Descriptions of data artifacts: PowerDesigner's metamodel contains all key fields needed to complete the standard descriptions of data dictionary items, including separate comments, rich text notes and annotations, editor and creator metadata, versioning information and more. For additional details not already provided by PowerDesigner, administrators can configure extended model definitions that provide extended attributes and custom forms to make user management of added metadata fields easy and natural.



Entity-attribute relationships: The Conceptual Data Model (CDM) is designed to have a list of data items independent of entity attribute participation. This can be taken advantage of in a number of ways: by disallowing reuse, users can be sure that a data item represents an entity attribute only once. Users can scan the model for data items that are not yet participating as an entity attribute. This exercise assists in developing a normalized model, and can also indicate where data concepts are not fully completed. Users can allow for reuse, allowing for the duplication of a data item concept on more than one entity as an entity attribute, either as a copy (two data items with the same name, but completely divergent definitions) or as a reused data item (one data item participating as an attribute equally to more than one entity, with one common name and one common definition). The choice of which method to follow largely depends on the definition of the corporate dictionary itself (is a 'name' a 'name', whether it is a customer name or a company name, or are all data items uniquely identified by their business name, and as such will not have a redundancy possibility). Whatever choice is made, PowerDesigner enforces consistent practice.



Entity to Entity Relationships: The CDM allows for easy definition of relationships between entities, supporting one-to-one, many-to-one and many-to-many relationships as well as supertype/subtype relationships using a number of industry standard notations including IDEF 1/x, IE, Merise and Barker's notation



A data dictionary tool must also provide for the following best practices

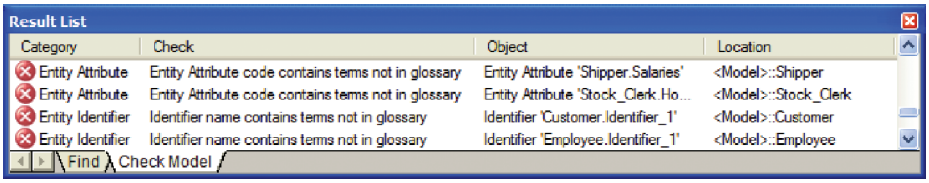
- Archived obsolete concepts based on the changing business
- Allow for data concept expansion
- Represent data in normalized models
- Support governance and standards enforcement

Archive obsolete concepts: In PowerDesigner, there are several ways to support the concept of obsolete data concepts. PowerDesigner's full traceability allows for identifying and preserving the history of the 'obsolete concepts'. When encountered again, the description of the 'obsolete concept' can quickly be uncovered and then a business decision is made about how to proceed.

Allow for data concept expansion (more attributes, more relationships, versioned): The enterprise repository allows for versioning of all model objects at the object level, and for comparison between any two versions of models and objects within them. As the concept expands, new details can be saved in a new version of the object, keeping the older versions around in case there are version-specific dependencies on the implementation side to keep track of. When a specific set of models needs to be used together at a specific version of the data dictionary model or models, the use of the repository configuration feature ensures proper version matching. By selecting a specific document/version for each document to be used together, PowerDesigner can be used to ensure strict pairing between dictionary and project when version "lock-in" or history is needed.

Represent data construct in a normalized model: The CDM allows for easy normalization, independent of logical and physical details. PowerDesigner is designed to view the conceptual data model as a level of abstraction above the storage paradigm and the implementation platform. The conceptual data model is not a relational model (does not have migrated foreign keys, allows for unresolved many-to-many relationships) and can therefore be normalized in a way that is consistent with the pure data in concept, not in implementation. In other words, all entity attributes of a given entity are there because they are functionally dependent on the entity identifier, and are not used in any way in any other entity. Foreign keys are artificial. For example, an "order" concept should not have a "customer identifier" attribute, as this attribute is not really functionally dependent on the "Order Identifier" itself, but is a construct inserted to implement the relationship (ie: a place to store the data that defines the implementation of this relationship concept within a relational database) and can confuse the analyst looking at "Order" for its pure data definition.

Provide metadata governance to enforce rules and practices to increase consistency of the dictionary as a standard: PowerDesigner provides a series of techniques to extend the built in modeling and metadata governance rules in a way that is seamless to the end user. Using these techniques, PowerDesigner can be used to extend the check model to include custom governance rules ensuring not just good dictionary development and maintenance, but standardized along your own set of criteria. Using events you can also ensure, in real time, that inaccurate or inconsistent metadata is caught at design time, instead of waiting for a report to reveal the gap. Using the passive or active enforcement is a matter of choice, whether it is better to force the user to take steps to correct gaps and mistakes up front, or allow for freer thinking workflow with a need to have the model checker assist in cleaning up the work after the fact.



For information on our comprehensive Consulting and Education Services to support your Sybase technology initiatives, visit us at www.sybase.com/consulting.

Sybase, Inc.
Worldwide Headquarters
One Sybase Drive
Dublin, CA 94568-7902
U.S.A
1 800 8 sybase

www.sybase.com

SYBASE® An SAP Company