Semantic Data Models

Introduction

Description: A semantic data model describes the concepts that are important to an organization along with their meanings and relationships to other important concepts. It is something like an authoritative Glossary of Terms along with diagrams or other ways of showing the relationships between different data. A semantic model has an emphasis on relationships and meaning and how the data relate to the real world. Ideally, the highest level semantic data model for an organization should be able to fit on one page, and provide the ability to drill down for more detail.

A semantic data model is not an entity relationship diagram; it is not a relational model; it is not a UML class diagram, although UML practitioners sometimes use class diagrams to illustrate data concepts.

Goals: Goals of creating a semantic data model are:

- to get agreement and clarity on the meanings of concepts that are important to an organization or a business domain.
- to identify the systems of record that contain these important concepts.
- to identify where there are different contextual definitions for the same concepts within the organization.
- to be able to provide the same definitions to end users whether they access data through client applications which are consuming standard APIs or through reports from the enterprise data warehouse.

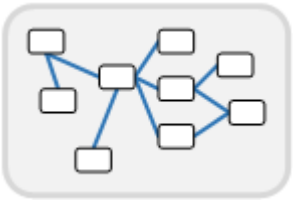
A semantic data model is used in building common understanding of things that are important to the organization in achieving its goals and objectives. It literally provides a common vocabulary.

A semantic data model focuses on the nouns and how they integrate with one another. It is very useful for building consensus and general understanding throughout an organization. Because it is a conceptual model, the scope is high level, but it can be scoped to any business domain.

Source: From the Wikipedia article on semantic data models:

> *The need for semantic data models was first recognized by the U.S. Air Force in the mid-1970s as a result of the Integrated Computer-Aided Manufacturing (ICAM) Program. The objective of this program was to increase manufacturing productivity through the systematic application of computer technology. The ICAM Program identified a need for better analysis and communication techniques for people involved in improving manufacturing productivity. As a result, the ICAM Program developed a series of techniques known as the IDEF (ICAM Definition) Methods which included the following:[1]*
>
> - *IDEF0 used to produce a "function model" which is a structured representation of the activities or processes within the environment or system.*
> - *IDEF1 used to produce an "information model" which represents the structure and semantics of information within the environment or system.*
>   - *IDEF1X is a semantic data modeling technique. It is used to produce a graphical information model which represents the structure and semantics of information within an environment or system. Use of this standard permits the construction of semantic data models which may serve to support the management of data as a resource, the integration of information systems, and the building of computer databases.*
> - *IDEF2 used to produce a "dynamics model" which represents the time varying behavioral characteristics of the environment or system.*
>
> *During the 1990s the application of semantic modelling techniques resulted in the semantic data models of the second kind. An example of such is the semantic data model that is standardised as ISO 15926-2 (2002), which is further developed into the semantic modelling language Gellish (2005). The definition of the Gellish language is documented in the form of a semantic data model. Gellish itself is a semantic modelling language, that can be used to create other semantic models. Those semantic models can be stored in Gellish Databases, being semantic databases.*

Scenarios



Links

- Semantic Data Model at the Gartner IT Glossary
- Semantic Data Models explainer by GoodData
- Semantic Data Model entry in the Encyclopedia of Database Systems (you might have access to this through your institution)

Contributors

Want to help with this page? Please see the Method Contributor Guide.

Stewards for this page:

- Leo Fernig, University of British Columbia
- Paul Schurr, University of Washington

Other contributors:

- Dana Miller, Miami University of Ohio
- Bob Dein, Miami (OH) University
- Troy Martin, BYU
- David Roberts, University of Michigan Medical School
- Scott Fullerton, University of Wisconsin Madison
- Jose Cedeno, Oregon State University
- Rupert Berk, University of Washington
- Robert Dean

A semantic data model can be used to serve many purposes. Some key objectives include:

- *Planning of Data Resources*: A preliminary data model can be used to provide an overall view of the data required to run an enterprise. The model can then be analyzed to identify and scope projects to build shared data resources.
- *Building of Shareable Databases*: A fully developed model can be used to define an application independent view of data which can be validated by users and then transformed into a physical database design for any of the various DBMS technologies. In addition to generating databases which are consistent and shareable, development costs can be drastically reduced through data modeling. Identifying business data that needs to be consistent and shared between different units. Having a single source of truth for central data that is at the core of various business units.
- *Evaluation of Vendor Software*: Since a data model actually represents the infrastructure of an organization, vendor software can be evaluated against a company's data model in order to identify possible inconsistencies between the infrastructure implied by the software and the way the company actually does business. Help identify areas where the design of the integration with 3rd party software should be done via loosely coupled APIs.
- *Integration of Existing Databases*: By defining the contents of existing databases with semantic data models, an integrated data definition can be derived. With the proper technology, the resulting conceptual schema can be used to control transaction processing in a distributed database environment. The U.S. Air Force Integrated Information Support System (I2S2) is an experimental development and demonstration of this type of technology applied to a heterogeneous DBMS environment.
- *Identifying Data Scope Elements*: Communicates about data boundaries, data ownership, and data redundancy.

Method

Skills:

- Ability to understand technical aspects of data/information
- Ability to understand the nouns that relate to business processes and capabilities.
- Ability to do data/information modeling.
- Ability to understand what is core data (entities) and what are important attributes to capture
- Language skills; ability to "tell a story"

Roles:

- Information Architect
- Enterprise Architect
- CIO
- Business users
- Others?

Steps:

- Identify the different elements of the solution
- Identify how the elements relate to each other

Tools:

- Can be as simple as a database and a web front end. The important part is wide visibility.
- Visio/PowerPoint
- Lucidchart

Communication

(to be completed)

Examples

UW-IT Investment Planning Objects and Definitions:

- https://itconnect.uw.edu/guides-by-topic/it-sourcing-standards/it-sourcing/it-investment-planning-approval/ in the new website at The University of Washington
- https://wiki.cac.washington.edu/display/EA/UW-IT+Investment+Planning+Objects+and+Definitions is the old-and-no-longer-accessible UW wiki site.

UW Financial System Glossary:

- https://f2.washington.edu/fm/myfd/home
- https://ucs.admin.uw.edu/myfd/Glossary.aspx

ITANA reference architecture for teaching and learning (RATL): https://spaces.at.internet2.edu/display/itana/Conceptual+data+model+v04

Note: We need a more canonical example to point to. UW is exploring using graph databases to illustrate the semantic data model. This should be a big improvement on a flat glossary of terms or a one page description.

Related Methods

Capability Maps: Semantic Data Models can be aligned and utilized in concert with Capability Maps to engage stakeholders.

Process Maps: Semantic Data Models can help explain how different processes are linked and what data are acted on and how.

Synonyms:

- canonical data model
- conceptual data model

Research (related links):

- http://en.wikipedia.org/wiki/Semantic_data_model

Note that these resources previously hosted at databaseanswers.org have been moved to https://datamodels.databases.biz/ and as of January 2024 are available there only partially. Those resources are available, also partially, at The Internet Archive ("IA" links included below).

- http://www.databaseanswers.org/data_models/semantic_building_blocks/facts.htm (IA)
- http://www.databaseanswers.org/data_models/canonical_data_models/canonical_and_semantic_models.htm