

DATA TOPOLOGY

By: Robert Conway, Director - Information Architecture Services, NEOS LLC
www.neosllc.com

Data Topology: A means of viewing the landscape of information assets available to a corporation in a discriminating fashion.

In the early days of data administration, a data model was described to the uninitiated as a representation of the business in terms of its data. A data model for IT was analogous to a chart of accounts for the comptroller, a floor plan for building maintenance, an organization chart for Personnel, a product catalogue for sales.

Today we live in the electronic world of commerce. From the members of the board of directors to their grandchildren in kindergarten, the means to gain immediate access to information is understood. And the notion that information is power has no dissenters. But the consumers of data are not a homogenous group. The domain of data in which individuals are interested is diverse. And even where the need exists for access to the same underlying data, the means by which it is assembled into information and the way that it is most effectively viewed and understood are seldom the same.

It is no longer sufficient to have an inventory of data. It is no longer sufficient to be able to provide a two-dimensional data representation of location and definition. Additional dimensions are needed. Filters are required. Points of reference are required. Different renderings for different purposes are required. Underlying data is no longer confined to traditional databases. Data is now a multi-media ingredient for information. To harness the power of information for the spectrum of potential consumers and their individual needs, we must map data as a topology and leverage a toolbox of facilities to support both data capture and data consumption services.

A **Data Topology** is a high level mapping of the key information assets of the business. It incorporates facts about location, relationships, quality, current and potential use, and classifications to assist with information management and strategy. Conventional maps take on different forms to effectively highlight different aspects of a region for a variety of audiences and for different purposes. Likewise, a data topology incorporates facts from multiple perspectives to serve different audiences and purposes. With a defined scope as its boundary, the key documents that comprise the data topology identify data flows among databases, the relationship of databases to products and markets, a mapping to business domains and high level data entities, an alignment of existing databases and databases under development to business operations and each other.

Key documents delivered as part of the Data Topology:

Topology Domain Hierarchy: An illustrative breakout of the Topology Domains to serve as a definition of the respective domains as well as a means of seeding analysis at a more granular level of detail. The Topology Domains are major information categories that capture the essence of how the business perceives the data encountered in performing its core functions.

System Genre: This document can be used to identify complimentary systems serving different markets in the same information or business function space, potential targets for retirement as new development projects are undertaken, sources for business rules and database documentation to support a subsequent detailed analysis of data and applications, the dependency of major business products or services on respective systems, and an inventory by market of business critical applications. High level business entities are used to classify each system. The entities are objects that are the focus of the corporation's core business as it relates to its products and services. These are the things about which data is captured or created.

Data Context: This document can be used to analyze the inter-system data flows and dependencies, gain a high-level understanding of the significant external sources of data at the corporation, gain an understanding of the

functional boundaries and touch points as different systems support various portions of the information lifecycle for one or several markets, gain a quick insight into data quality considerations that would have an impact on the business priority for system replacement/refurbishment or the choice of a database as a system of record.

Technical Information: This document can be used to locate where the data physically resides to support activities such as data extraction for test data and data quality analysis, network implications for inbound and outbound data, hardware & licensing impact of application retirement. On a practical level, it can provide a clue about who needs to be contacted for information on back-up and recovery, operational outages, disaster contingency plans and other aspects related to physical location and custodianship.

Interview Notes: These provide specific details provided by the contacts identified as knowledgeable about the respective capabilities and challenges for one or more of the systems targeted for analysis.

Determining the scope of the data topology will directly drive the topology domains and list of contacts to be included. A Data Topology can be developed incrementally with successive functions, locations, market segments, or technologies used as a basis of expansion. The following is an example of a typical starting point for an initial iteration:

The following items were considered "in scope" in binding the Enterprise Information Topology effort:

- U.S. based information systems maintained by IT corporate services
- Systems running on a mid-range or higher servers
- Systems which produce, or otherwise directly support, the corporation's products
- Business entities to be defined are bound by the identification of the entity and its relationship to others

Using these criteria, 21 systems were identified as the targets for analysis and 9 topology domains were included.

With an iteration of the key documents completed, two **companion documents** should then be developed: the Strawman Data Model and the Data Lens Classes.

A **Strawman Data Model** is a high level conceptual data model and provides a business representation of data within the corporation. This representation, in conjunction with the Data Topology, provides a framework by which new system applications can be designed and/or sourced. It provides a common starting point for all applications that require a database, as the business relationships should be consistent and comprehensive. Given the significant investment represented by application development initiatives, and high expectation that all applications will perform in harmony upon completion, it is vital to have a unifying point of reference. The minimum set of deliverables should include:

- High-level entity-relationship diagram within each of the topology domains outlined in the Data Topology
- Definitions of all entities and key relationships
- List of nominated /certified key reference data if appropriate

The purposes of the Strawman model as a key design deliverable are:

- Depicts the scope of core corporate activity
- Defines how entities within that scope relate to each other
- Creates a common model for understanding how business activities support or depend upon each other

Data Lenses are a means of viewing the landscape of a Data Topology. It is the filter that highlights information of interest to a specific audience and hides other portions of the Topology Domains. With the clutter removed, the lens can be mounted on the proper device for viewing. These devices might be dashboards, visualization software, data quality profile reports, or business information (BI) tools. The Data Lens is ground as appropriate to the individual viewer. In the information-intensive financial services industry, the following five lenses are typically encountered: Data Acquisition, Data Manufacturing, Data Publication, Customer Relationship Management, and Reference & Roll-

Up. In Retail operations, these lenses are more likely: Inventory Acquisition, Order Management, Shipping & Distribution, Customer Relationship Management, and Reference & Roll-Up.

Process to Create a Data Topology

- Start with discovery of the business. The core functions and competencies of a business according to the way an organization perceives itself become the Topology Domains.
- Identify the major systems that support the business functions and a subject matter expert (SME) for each system. Conduct the topology interviews to gather information about supported products, data flows in and out of the underlying databases, location of documentation for those databases, and any known data quality issues.
- Profile the databases and construct a Strawman Data Model. Develop context diagrams that depict data flows. Relate each database to the applicable Topology Domain and Strawman entities.
- Present the findings to the respective SME's to discover/confirm additional insight regarding data quality, data flows, lineage and relationships. If any information items for the Topology template are outstanding, request assistance from the SME.
- Document the hierarchies and subtypes within each Topology Domain to add clarity and definition. Use this additional level of detail to develop a more detailed Strawman with ER diagrams for each Topology Domain.
- Using the business requirement (motivation, challenge, IT initiatives, business programs) discovered throughout the preceding activities, define a set of Data Lenses that represent relevant views of the Data Topology information. Document domain views for each of the Data Lenses.
- Evaluate the databases for respective potential as a data book of record. Use quadrant diagrams depicting business scope, extensibility, and origination to capture the result of this assessment.
- Present findings and recommendations to the project sponsor and target audience.