# TABLE OF CONTENTS

1. The Absence of Data Governance Threatens Business Success
2. Data Repurposing and Data Integration
   2.1 Example: Structural Modeling Precision at the Data Element Level
   2.2 Example: Structural Modeling Precision at the Table/Relationship Level
   2.3 Example: Semantic Differences
3. Challenges for Centralization and Repurposing
4. Data Architecture and Data Governance
5. Data Modeling and Architecture Standards
6. Maintaining Relevant Enterprise Metadata
7. Data Requirements Analysis
8. Considerations: Effectively Governing Data Architecture
9. About the Author
10. About Sybase PowerDesigner®
THE ABSENCE OF DATA GOVERNANCE THREATENS BUSINESS SUCCESS

The absence of data governance standards is a critical failure point for enterprise data repurposing. Ineffective oversight of data representations, semantics, and models introduces severe risks to “silver bullet” applications such as Enterprise Resource Planning (ERP) or Customer Relationship Management (CRM). Business applications developed in a virtual vacuum have ill-defined business terms and a variety of models for common data concepts, creating an aura of suspicion around consolidated data sets. Yet the exploding demand for data repurposing poses recurring questions about the trustworthiness of data managed in an enterprise data warehouse or master data repository.

This paper examines the challenges associated with the root causes of data centralization failure, and then reviews straightforward best practices typically ignored when systems are designed in an ad hoc, organic manner (as in most organizations). Instituting these data governance best practices will reduce the risks and increase trust in organizational information:

- Data architecture and data modeling standards to reduce variation in structure;
- Enterprise metadata management (from a horizontal perspective) to standardize semantics and to provide visibility of use from concept to instantiation; and
- Comprehensive data requirements analysis to capture all prospective data consumer requirements.

Implementing these best practices requires the integration of processes and technology, specifically data requirements management, metadata management, and data modeling. However, these tools are employed most effectively when knowledge captured within any part of the technology can be shared across the entire application development lifecycle. When the tools and techniques provide a line-of-sight during the design phases from the requirements through to the implementation and transition into production, a link can be made from concept to data instance so that all system impacts can be identified for any adjustments of changes in semantics or structure at all levels of data precision.

DATA REPURPOSING AND DATA INTEGRATION

In the early days of computing application programs and their underlying data sets were developed to address specific business application needs within specific areas of the organization. The Sales Department applications differed from those developed to support Fulfillment, and likewise for the application supporting Finance, Back Office processing, and a number of other lines of the business.

In contrast, today there is a growing trend of data repurposing, in which selected business applications discover and ultimately reuse data sets created or acquired to meet one business application’s requirements for alternate purposes. We are familiar with the grand, “enterprise-level” examples — Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), business intelligence and analytics, and even reporting via data warehousing, which all rely on data integrated and consolidated from across a collection of source applications. In turn, business consumers of many additional applications expect to benefit from the unified views of common business concept areas incorporated and managed within master data environments.

Yet the expansion of the scope of use of repurposed data sets exposes challenges and conflicts that can potentially wreak havoc on the intended results of the consuming applications. For the most part, siloed data models and applications have been designed in a vacuum, with little concern for interoperability across the line-of-business boundary. The data sets were mostly developed to support specific transactional or operational needs, and therefore they have been engineered to satisfy immediate requirements without any consideration for longer-term downstream consumption.
And although the same business terms had been used, the absence of rigor in enforcing naming standards or providing clear definitions meant that differences in structure, format, and meaning have crept into the data. When data sets are used for their original purpose, these variances in structure and semantics are largely irrelevant. But the byproduct of data repurposing is the magnification of these structural and semantic differences. The result is that ungoverned consolidation will expose increasing complexity and difficulty in successful reuse of data for alternate purposes such as Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and Master Data Management (MDM).

**CHALLENGES FOR CENTRALIZATION AND REPURPOSING**

The desire to repurpose data must be contrasted with the challenges in effectively centralizing the data sets to support the cumulative business needs. With data sets that are under consideration for centralization, even slight structural and semantic variances can inadvertently introduce inconsistencies for downstream consumers, especially after a series of data transformations are applied to force data sets to merge into a (often hastily-engineered) target representation.

Although incomplete attributes or variance in which values are perceived to be accurate can be the culprits, more often the issues of inconsistency emerge as the byproduct of the absence of historical standards (and lack of governance) for the ways that different stakeholders model their core data concepts. So while organically-developed applications are likely to share representations of the same concepts, their “silos” development often leads to structural differences at various levels of precision (e.g., data element vs. table structure), as well as semantic differences at the many levels of precision.

**Example: Structural Modeling Precision at the Data Element Level**

Even commonly-used attributes are subject to structural variation. Consider a data concept with a well-defined standard — the North American Numbering Plan, which is the standard for representing telephone numbers, used in the United States, Canada, and a number of other countries. The standard specifies a telephone number structure “+1-NPA-NXX-xxxx,” in which the “NPA” refers to an area code, the “NXX” is a central office exchange code, and the “xxxx” is the subscriber (or line) number.

However, consider the many ways that telephone numbers are presented, using a variety of special characters (including parentheses, hyphens, periods, commas, and spaces). The underlying data elements are structured many different ways; just a few examples are shown in Table 1.

<table>
<thead>
<tr>
<th>DATA ELEMENT REPRESENTATION</th>
<th>EXAMPLE</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAR(12)</td>
<td>“301-754-6350”</td>
<td>NPA, NXX, subscriber number all separated by hyphens</td>
</tr>
<tr>
<td>CHAR(10)</td>
<td>“3017546350”</td>
<td>All punctuation removed</td>
</tr>
<tr>
<td>NUMERIC(10)</td>
<td>3017546350</td>
<td>Numeric representation</td>
</tr>
<tr>
<td>VARCHAR(15)</td>
<td>“+1-301-754-6350”</td>
<td>Allows for prefixed “+1-”</td>
</tr>
<tr>
<td>VARCHAR(20)</td>
<td>“(301) 754-6350 x101”</td>
<td>Allows extensions, alternate representations</td>
</tr>
</tbody>
</table>

Table 1: Example data element representations for a NANP telephone number.

Even these few examples demonstrate differences that require parsing, standardization, and resolution of structure (especially when there are embedded extension numbers!) when attempting to repurpose data sets. And this is but one example using data values that are already subject to existing standards — consider the challenges with data elements whose values are not expected to conform to a defined standard.
Example: Structural Modeling Precision at the Table/Relationship Level

As a means for establishing contact, telephone numbers factor in structural modeling precision issues as well. Early data tables and files used for batch customer transaction processing may have been designed to capture one, and possibly two telephone numbers — the customer’s home telephone number and possibly an office telephone number. But files structured with column space to hold only two numbers cannot capture the many possible telephone numbers that today could be associated with an individual, including mobile numbers, Voice over IP (VOIP) numbers, virtual office numbers, fax numbers, as well as many other contact mechanisms. Later system designers will have dissociated the data attributes associated with contact mechanisms into related tables linked via foreign keys. These structural differences introduce the need for more complex rules and transformations in order to reuse data from different sources.

Example: Semantic Differences

The potential complexity of the structural variation is dwarfed by the challenges of variant semantics. Just consider the many meanings for commonly-used business terms. For example, to the sales organization, a customer is a party who exchanges value in return for products or services. But to the customer service organization, a customer is a party entitled to customer support services. In the situation where evaluation products are provided at no charge to interested prospects, there is a qualitative difference between “sales customers” and “service customers,” even though they are both referred to as customers.

DATA ARCHITECTURE AND DATA GOVERNANCE

The challenges introduced by the absence of governance in legacy systems designs means that the growing interest in repurposing data from across (and even from outside) the enterprise suggests that moving forward, modeling and metadata management cannot be performed in a vacuum. Rather, oversight at the organizational level must be imposed to establish standard practices for enterprise data design, modeling, sharing, and reuse. This suggests the need for specific policies for data governance associated with different aspects of data architecture, with the intention of establishing a high level of maturity and capability, namely:

- Data architecture and data modeling standards to reduce variation in structure;
- Enterprise metadata management (from a horizontal perspective) to standardize semantics and to provide visibility of use from concept to instantiation; and
- Comprehensive data requirements analysis to capture all prospective data consumer requirements.

DATA MODELING AND ARCHITECTURE STANDARDS

The high likelihood for data reuse will influence system developers to approach their designs in a way that anticipates downstream data consumption. Standardizing the representative models will reduce the effort for subsequent extraction and consolidation, and this means increased oversight over any newly-developed data models. Instituting organizational standards along with the data governance processes overseeing observance of these standards is the first step to resolving the challenges inherent in wholesale data consolidation.

Governing areas of data architecture combines defining policies for observing data element standards, data modeling guidelines, coupled with the processes to ensure that those standards are observed. This may run the gamut from rudimentary policies defining data element naming conventions, normalizing structures for common data themes, defining schemas and canonical models for data exchange, to establishing protocols for enterprise data modeling as well as instituting processes for data model review and acceptance by the members of a data governance board.
MAINTAINING RELEVANT ENTERPRISE METADATA

The flip side of defining organizational data element and modeling standards involves communicating the details of, and then managing compliance with those standards. One effective method to accomplish both of these goals uses metadata management methods. And when the data management practitioners within the organization understand the ramifications of slight variations, they strive to attain a high level of metadata maturity.

This means that a metadata management strategy is clearly defined and communicated to all developers and consumers, and there are centralized tools and techniques integrated as part of the enterprise development framework. A single metadata repository accessible across the organization can be used to document data element concepts, their instantiations, and any structural variances. Business terms can be mapped to data element concepts, which are then linked to their assorted instantiations across the application infrastructure. This provides a virtual “line-of-sight” between business concept and application use. Where the conceptual data elements are touched by more than one business application, the metadata analysts can review the usage map for those elements and analyze the impact of adjustments to any underlying or dependent data element definitions.

DATA REQUIREMENTS ANALYSIS

We are conditioned to consider the business application that either creates or initially acquires the data as the “primary consumer.” But while primary use of a data set can be defined as “first in order,” it can also be defined as “first or highest in rank of importance.” Increased data repurposing means that the originating application may not be the most important use of the data. If the alternate uses are high in rank of importance, they are also primary consumers. Therefore, it is critical to ensure that measured levels of structural and semantic consistency are sufficient to meet the business needs of the collected downstream data consumers, which means thinking differently about soliciting and documenting data requirements across the organization.

Usually, data requirements are a byproduct of the functional requirements implied by the needs of the business process whose application is being designed; in turn, those data requirements are only defined to meet an acute functional need, but do not address how the data sets are potentially used by other business processes. But as more data sets are subjected to centralization and repurposing, there is a corresponding need to adjust the system development process so that enterprise requirements are properly captured and incorporated into the data architecture.

Yet again, data governance policies can help direct an approach to soliciting, capturing, and documenting data requirements in a way that can be directly linked to the ways the underlying models will be designed and implemented. Guiding the ways that system designers engage the general community of potential data consumers will ensure that organizational information requirements are captured and managed. This reduces the need for downstream data extractions and transformations while improving general information usability. Instituting good data quality practices and governing those practices with the right tools and techniques essentially reduces structural and semantic inconsistency.
CONSIDERATIONS: EFFECTIVELY GOVERNING DATA ARCHITECTURE

As the rates of data volume growth continue to rapidly increase, our technical advisors suggest that unifying our views of enterprise information via enterprise data warehouses, enterprise resource planning, or master data management will increase value along the different value driver dimensions such as increased growth, decreased expenditures, and reduction in risk. Yet we have shown that the traditional approach to data repurposing, consolidation, and reuse itself entails a number of intrinsic risks. To avoid these risks, you must institute data governance.

Yet defining selected data governance policies only addresses one piece of the puzzle. When data policies are defined, there must be processes and procedures to back them up, with corresponding methods to develop business applications while meeting business objectives within a data governance framework. This suggests considering tools and techniques to oversee approaches to organizational data architecture that support enterprise information management and governance goals, including:

• Processes for defining and then approving data policies;
• Communication of data policies and associated guidance across line-of-business boundaries;
• Unifying the collections of data requirements for all key data concepts;
• Documenting structural data standards;
• Harmonizing business term definitions and semantics;
• Unified business models and developing standard data models for persistence and sharing;
• Active monitoring of observance to data expectations, and
• Assessing the requirements from across the line of business landscape and ensuring consistent observance of those requirements through the design, development, and implementation phases of the system.

Even if data governance practices are defined, there must data management tools and techniques to ensure a line-of-sight during the design phases from the requirements through to the implementation and transition into production, including data requirements management, metadata management, and data modeling.

More importantly, though, these tools must support the sharing and exchange of knowledge throughout the development lifecycle. Data expectations captured during the requirements gathering stage must be connected to the associated data elements and data models that are used by the developed business application. At the same time, during all system development life cycle phases, the data requirements must remain visible, maintaining the link from concept to data instance so that all system impacts can be identified for any changes in definition or structure at any level of data precision.

While disparate tools may support some canonical representation for sharing metadata, attempting to cobble these tools together may not only introduce additional system development complexity, it may actually lead to a chaotic environment in which much of the data governance practices are wasted. Employing data management tools that are inherently engineered to provide visibility both across the data architecture and along the system life cycle effectively supports the integration of data governance policies and practices to enhance data reuse across the enterprise. These become the critical success criteria when evaluating tools to support governed data management.
ABOUT THE AUTHOR
David Loshin, president of Knowledge Integrity, Inc. (www.knowledge-integrity.com), is a recognized thought leader and expert consultant in the areas of data quality, master data management, and business intelligence. David is a prolific author regarding BI best practices, via the expert channel at www.b-eye-network.com and numerous books and papers on BI and data quality. David can be reached at loshin@knowledge-integrity.com.

ABOUT SYBASE POWERDESIGNER
PowerDesigner, the industry leading modeling and metadata management tool, offers a model-driven approach to empower and align Business and IT. Implementing data governance best practices requires a tool that allows all levels of data to be captured, articulated and shared. Today’s complex IT environments require tools and processes to manage the flow of information between all phases of the IT organization and business community.

PowerDesigner’s true impact analysis reduces time, risk and cost associated with changes within the BI environment by:

- Establishing a single version of the “truth” for key information assets
- Providing consistent information, when and where needed, to improve decision making
- Enforcing governance and accountability for key information assets in support of compliance
- Enabling information to be shared and exchanged, with appropriate safeguards
- Supporting efficiency, collaboration and transparency needs

PowerDesigner provides an integrated modeling solution that encompasses:

- Data Modeling
- Business Process Modeling
- Applications Modeling
- Business Requirements Modeling
- Metadata management
- Support for Enterprise Architecture frameworks