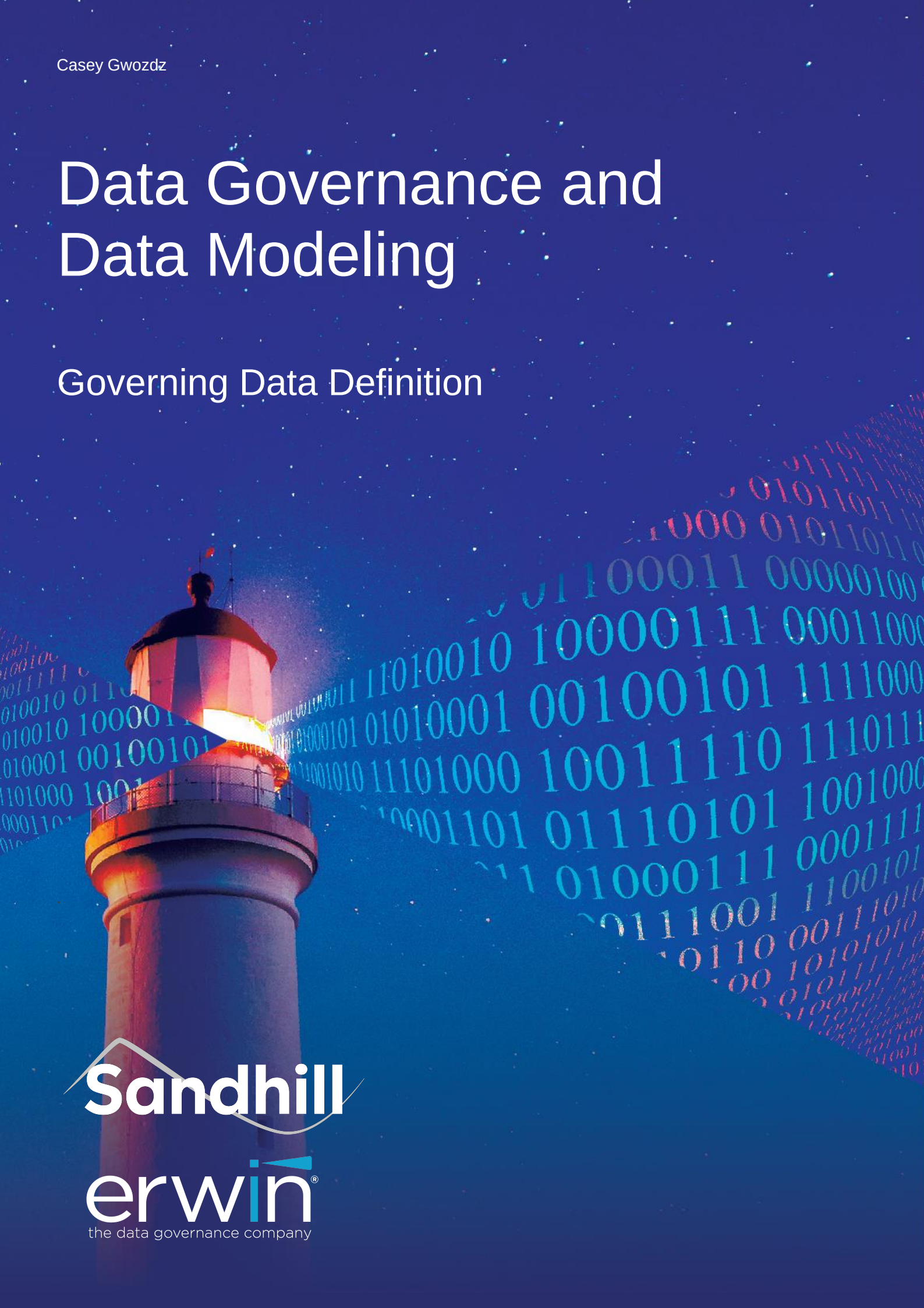
Casey Gwozdz

Data Governance and Data Modeling

Governing Data Definition



Sandhill

erwin®

the data governance company

SECTION 1

Introduction

Enabling understanding and collaboration through a common vocabulary.

INTRODUCTION

Many years ago, a school teacher named Robert Cawdrey had a very interesting idea. In his society, there was an explosion of international commerce and many new words were being developed for the English language. Also, literacy among English speaking people was becoming more and more widespread. It was becoming quite a challenge to help folks understand and use the English language. Robert Cawdrey thought to put together a list of words in a specific format with a list of abbreviations signifying the word origin along with a definition for the word to help promote understanding and word usage among English speaking people. So, after years of effort, Robert Cawdrey is credited with publishing the first dictionary of the English language in 1604, entitled "A Table Alphabeticall" (at least according to the internet). Since then, virtually all languages provide a standardized format for explaining and using words in their respective languages to aid in understanding and intended meaning in order to increase better communication with each other.

Skip forward a few centuries, and it seems we have a similar challenge in today's corporate computerized application business system society. More and more employees are becoming "computer savvy" with access to tools related to e-mail communications, spreadsheets, documents, etc. These tools even have dictionaries imbedded for spell checking and grammar review to help improve communication. However, when it comes to developing an understanding of the vast amounts of data stored in our numerous business application system investments today, employees really have a challenge. They still need to rely on static "documentation" (that can become stale after the first system enhancement), time in that position (experience), self-investigation, peer-to-peer learning, or maybe even a quick e-mail announcing specific changes or enhancements that have been applied to the business application system. As a result, learning to really understand the data supporting these business application systems takes time – lots of it!

In reality, as a number of publications have already stated, stored data values populated to support a business application system is meaningless. These data values are captured and retained in persistent storage (on disk in various data formats) in order to enable the application system to "know and remember" data items that support the business process and be able to present that data back to employees. Without some sort of indication what the populated data value represents (at least a report column header name or screen reference label) it is impossible to grasp what these data values truly represent. However, most of the column headers or screen labels employees get to see in a business application can be cryptic, misleading or do not fully clarify the actual meaning of these populated data values. This extends the learning curve for employees to look at a data value, understand what it is (information), learn to correlate that value to other data values (knowledge), and finally grasp enough understanding to identify process improvement or change and efficiencies (wisdom) within the context of the business application system, as well as across integrated business application system investments. Any business application system stored data value in this sense becomes a "data" asset. It has value to the company, its value diminishes over time, and can be deleted when it no longer provides value to the company.

All of this value and understanding about these corporate data assets wind up in the mind of those employees closest to and within the context of the application systems they are working with. Unfortunately, in today's human resource climate, employees move from position to position, leave the company, or, due to other corporate circumstances, are laid off. Whichever the case, all the information, knowledge and wisdom gained through an employee's experience related to these data assets is lost when the employee moves on.

Also, with new technologies now available, more and more business application systems are being deployed to mobile devices and running these applications are placed in the hands of the public outside the corporate domain. As a result, there are fewer and fewer business employees gaining the experience within the application process behavior to understand what is being stored in these application systems. This increases the necessity to define and understand the stored data values being populated in these systems in order to capture, develop and report stored application system information within the corporation.

The challenge then is: How can we leverage this human understanding about corporate data assets and consistently share them with all employees? A system that can support this can provide a method to quickly gain this understanding, contribute to process improvement, increase quality, identify opportunities to eliminate data redundancy, and come up with other great ground-breaking ideas (i.e. SAVE THE COMPANY MONEY!)?

SECTION 2

The Data Governance Practice

More than rules and control, Data Governance is about enhancing an organizations ability to leverage data.

THE DATA GOVERNANCE PRACTICE

From this perspective, a basic definition for a Data Governance initiative is to find ways to record, manage, track and maintain business application systems that can respond to this kind of challenge. Without addressing the details around the organization, personnel, system support, data storage requirements, and vendor product considerations involved in such an investment, the objective of a Data Governance initiative is to capture all the information, knowledge, and wisdom acquired and stored in our employee's minds (or not even thought of yet) about our stored data assets. This can provide the vehicle to enable sharing this valuable resource across the organization to encourage business improvement within our complex computerized business application system society.

Of course, there is lots of additional "stored data" that needs to be recorded for such an application to "know and remember" what is captured over time in an employee's mind through experience in working with a specific business application system. The term coined for this kind of data asset is "metadata" or data describing a data value that is stored in a business process application system. This metadata is needed to facilitate both business and technical understanding of a data values' lifecycle within the context of individual application systems. Furthermore, metadata can be used to link and summarize this understanding into a consolidated "dictionary" to encourage the spread of knowledge and collaboration across the ever changing set of employees that support and use our application system investments. A standard format has evolved for this kind of information and is referred to as a "Business Glossary."

The International Organization for Standardization (ISO) published standard # 11179 to establish fundamental ideas around support of these metadata requirements. This is somewhat akin to Mr. Cawdrey's idea for formulating a standard format for his dictionary to help better communicate and provide understanding around the definition, use, and application of words in the English language.

As an example, consider the annual report a public organization publishes primarily to inform their corporate owners about the financial results of the company. This report includes a number of financial statements that invariably include numerous footnotes that further describe and justify how these aggregate figures are derived or applied within standard accounting practices. In reality, these footnotes are indeed "metadata" that further communicates, clarifies, and assists the company owners in understanding the company financial results. Shouldn't these same standards be applied to the nature of valuable stored data assets across our many complex business application system investments? By providing the means for any employee to contribute new understanding related to one or more stored data items within the context of a specific application system, data governance personnel can review these contributions (i.e., comments) and consolidate them to a revised, clear, concise understanding in a "Business Glossary" as "the dictionary" supporting the computerized business application system society.

SECTION 3

The Relational Data Modeling Practice

Rigorous standardization reduces the risk.

THE RELATIONAL DATA MODELING PRACTICE

The relational data modeling practice is based on a rigorous methodology supporting a "new" technology break through some 4+ decades ago – Relational Database Technology. Relational data modeling tools came into the marketplace to support the underlying science behind the relational database technology paradigm. It initially became popular with technical developers in creating and supporting the best performing relational structures to support the data access needs of the front end business application systems.

In addition to supporting the technical database implementation directly from the data modeling tool, some tools placed a focus on augmenting the tool to store metadata from a "business" perspective to help non-technical personnel understand the data that is being stored. This resulted in an underlying "metadata model" that can support both the technical database structure as well as the data understanding from a business perspective.

erwin® Data Modeler was one of the first relational data modeling tools on the market. This product has a very mature model for storing metadata as it has supported metadata requirements for both business and technical audiences since its debut some 25 years ago. And, erwin Data Modeler has a very long history as the leading data modeling tool on the market.

Of course, the relational data modeling practice has different levels of maturity as implemented in different organizations.

Some organizations are content with using data modeling tools purely from a technical, physical database implementation perspective as shown in the fourth tier building blocks in Figure 1. Practices in other organizations incorporate populating the business understanding of stored data elements within the context of the application usage during the course of executing Systems Development Life Cycles (Figure 1, tier three). Still others incorporate integrating conceptual, logical (business) and physical data models in efforts to approach enterprise data modeling principals across multiple business application systems. Finally, organizations can store metadata within a data model to incorporate defining how one business application system "pushes" their stored data values to another business application system (i.e. Master Data Management hubs, Data Warehousing, etc.). Other terms for this process are "data mapping," "Extract, Transform, Load (ETL)," or "Source-to-Target Mapping." These approaches are the beginnings of Figure 1, tier three. erwin Data Modeler has a long history of supporting all of these data modeling practices for a number of years.

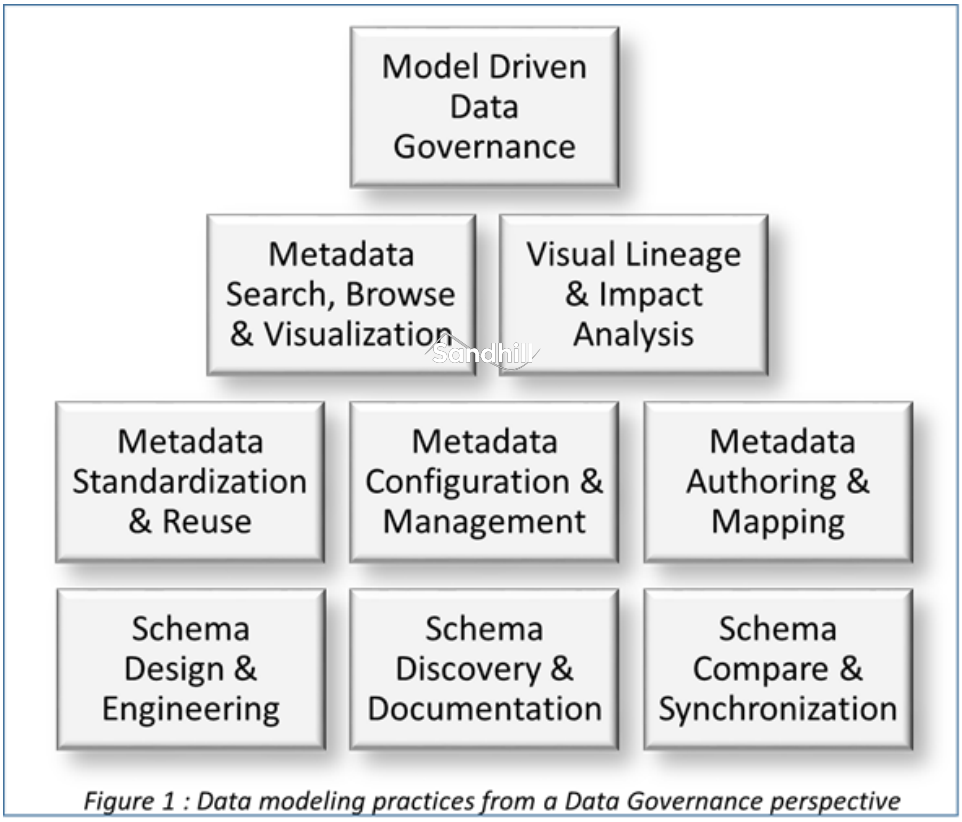


Figure 1 : Data modeling practices from a Data Governance perspective

There are also a number of other vendor products that include stored metadata (i.e. BI tools, ETL tools, UML tools, etc.) for employee easy reference while using these tools. Employees can view the metadata about the business application stored data values they are viewing within the context of the business application system within these products for better understanding and appropriate usage. erwin Data Modeler provides the vehicle to import and/or export the valuable metadata related to the stored data structure in these vendor products to guarantee the same metadata co-exists in multiple vendor products and employees are looking at the same information about the stored data values across vendor products. Simply put, if the data model can become the "record of source" for both business and technical metadata information, this information can easily be replicated to these other vendor products so that all employees are viewing same version of the truth.

erwin Data Modeler includes more than 120 import/export bridges for integration to these vendor products that include metadata information. This also includes a number of Big Data distributions used to flush out business understanding around these analysis investments and/or to discover possible integration points to existing application system investments using relational technology.

In addition, there are a number of "ERP" packages available in the marketplace that provide business application systems in lieu of building the system in-house. These vendor products include very complex stored data structures to support these application systems. Also included are vast amounts of metadata about these stored data values. The erwin Safyr Option offers a process to manage and extract this metadata into the data modeling practice to, among other functions, easily replicate this metadata into the other vendor products mentioned above for consistent messaging related to ERP application system data.

SECTION 4

Data Governance Supported by a Relational Data Modeling Practice

Building a solid and reliable foundation for data governance, regardless of the scale, scope and goals.

DATA GOVERNANCE SUPPORTED BY A RELATIONAL DATA MODELING PRACTICE

So, the erwin Modeling product team had an idea. Why not leverage the 25+ years of working with a metadata framework that supports both business and technical understanding of stored data as part of a Data Governance practice? After all, a data model is the closest and most current reflection of the actual physical data store supporting each business application system investment and can also provide the business metadata understanding within the context of these business application system investments that employees need to deal with every day

The first new product delivered several years ago is called the erwin® Web Portal and it was made available in two editions: "Standard" and "Enterprise." This product can import or "harvest" erwin Data Modeler data models into a web based environment. Users can package together sets of data models that represent the live physical data stores that support each business application system and manage the viewer access based on pre-defined audiences (both business and technical). All the metadata stored in erwin Data Modeler data models is now accessible to wide audiences. If leveraging data modeling practices using data mapping capabilities in erwin Data Modeler, this product can display both graphical and textual data mapping characteristics down to the populated data level to support tracing data impact analysis (if a source application data item is changed, what is the impact downstream?), data linage analysis (How is this data item derived from upstream application systems?), as well as model to model semantic lineage (How and where is this physical data item related to a business logical model and/or a conceptual model perspective and what does it mean?).

In March 2015, the erwin Modeling product team delivered an enhanced edition of the erwin Web Portal called the "Data Governance" edition. This edition added in an ISO 11179 standard business glossary where glossary terms can be directly related to the stored data items present in a data model. The data modeling practice maturity level currently in place is not as relevant to begin the process of achieving progress toward a data governance practice related to metadata management and data architecture, as well as other practices critical to Data Governance.

If the data modeling practice is currently at the physical only level, these physical stored data items can be related directly to business glossary terms via a standard workflow (create, review, approve and publish) process. Doing this promotes sharing, understanding, and collaboration based on the true meaning of data items stored in our business application systems for both business and business application system support (technical) employees.

If the data modeling practice is leveraging populating business metadata within the data modeling practice, these definitions can provide initial meaning and understanding of stored data values within the context of each business application system to assist data governance personnel to establish sound, concise, and meaningful business glossary terms. These terms provide a framework to best develop an understanding of stored data values across business application system investments.

"Appendix A: Data Modeling Practices supporting a Data Governance Initiative" provides a list of some of the data modeling practices supporting a data governance initiative leveraging the erwin Data Modeler product.

SECTION 5

Summary

To deliver real value, you need to look after the basics.

SUMMARY

A Data Governance practice involves strong collaboration among both business and IT technical employees. Of the many considerations involved with this effort, combining this practice with traditional data modeling practices can support many of the data governance activities related to Metadata Management, Data Management, Reference Data Management, and Data Architecture, just to name a few.

Any employee with access to the erwin Web Portal can offer ideas, comments, or enhancements to the understanding of a stored data value either within the context of a specific business application system (a data model), or, within business data glossary terms. Data Governance personnel can then in turn review these suggestions and follow up to incorporate this understanding into the best understanding of the stored data value for the benefit of entire company and its current employees.

Through a Data Governance practice, with supporting practices around data modeling, we can aspire to the ideas and objectives that Mr. Cawdrey had so many years ago to share in an understanding of data values being stored and maintained across our entire complex business application system investments and ultimately reduce company costs through business process improvement. Here is a reference from the preface of Mr. Cawdrey's first dictionary of the English language (using old English):

> ***A table alphabeticall, or the English expositor |h [electronic resource] : |b containing and teaching the true writing and vnderstanding of hard vsuall English words, borrowed from the Hebrew, Greeke, Latine, or French, &c. With the interpretation thereof by plaine English words, gathered for the benefit and helpe of all vnskilfull persons. Whereby they may the more easily and better vnderstand many hard English words, which they shall heare or read in Scriptures, sermons, or else where, and also be made able to vse the same aptly themselues. Set forth by R.C. and newly corrected, and with the adition of many vsefull wordes enriched.***

A data modeling practice is a key component within any data governance initiative. The erwin Modeling product line already provides the vehicle of consolidating business and technical metadata from various sources into a single version of the truth. Using erwin Web Portal Data Governance edition, an organization can enhance and extend those base capabilities provided to include a number of additional key capabilities related to a formal Data Governance practice.

ABOUT THE AUTHOR

Casey Gwozdz, Principal Consultant at erwin, Inc., has been involved with IT systems development and support for over 35 years with an emphasis on various modeling methodologies and practices centered on Batch Processing, OLTP processing, Process analysis, and Data analysis. His primary focus has been in helping customers leverage the erwin Data Modeling product line for the last 12 years.

APPENDIX A: DATA MODELING PRACTICES SUPPORTING A DATA GOVERNANCE INITIATIVE

erwin Data Modeler Option	Option Type	Impact supporting a Data Governance initiative
Model Theme	Modeling Object	Model Themes are erwin Data Modeler objects that support developing a standard look and feel for your data model diagrams. This includes font size, color, object shape frames, etc. Standardizing your model diagrams across all data models can help non-modeling viewers better understand diagrams when viewing them in the erwin Web Portal. Multiple themes can be supported in a erwin Data Modeler data model. You can set all diagrams to use a default theme by specifying a theme in the model root level in the Model Properties dialog. This default theme can be overridden at the Subject Area, individual diagram or even at the diagram object level to accommodate various needs. This provides a standard look and feel when viewing these diagrams across the business or non-data modeling user base in an integrated product like the erwin Web Portal Data Governance edition.
Domain	Modeling Object	In erwin Data Modeler, a domain object is used to identify a category name that includes other metadata to support consistency across attributes or columns in a data model. Business attributes as well as physical column objects can be associated to a domain object to "inherit" all the metadata properties defined. Flexibility is available to "override" the inheritance of the individual properties where justified. These domains can then be leveraged when the model is harvested to the erwin Web Portal to establish or associate these objects to domains in a Business Glossary.
User Defined Property (UDP)	Modeling Object	User Defined Properties (UDPs) are flexible ways in which a data modeler can add new attributes to virtually any erwin Data Modeler object. These UDPs can be used to populate pertinent metadata values to a data modeling object above and beyond the metadata properties available in the standard out-of-the-box metadata storage capabilities in erwin Data Modeler. These metadata labels and their populated metadata values can be viewed in the erwin Web Portal and assist in understanding more about the object perspective. Standards around UDP metadata property names can assist Web Portal viewers in understanding the model object from a data governance perspective.
Naming Standards	Modeling Object	Naming Standards is an object in erwin Data Modeler that enables you to translate business names from the logical or business view of a data model into abbreviations to the physical side of a data model. A glossary of terms are added to establish an abbreviation to be used as the corresponding physical object word in the physical implementation of the data model. There are also methods that support retrofitting existing database investments to translate schema object names to a more suitable business name from a business perspective. From a data governance perspective, when developing a business data glossary, data governance personnel can review the definitions as stated within the context of the application system to formulate the definition of the glossary term. In addition, all physical schema object names can easily be tracked as alias names.

Active Model Template	Modeling Function	An Active Model Template is a erwin Data Modeler data model used to establish specific data modeling standards for all data models in your environment. Key data modeling metadata populated in an Active Model Template (i.e. Themes, Domains, User Defined Properties, Naming Standards, etc.) can be pre-established as the standard set up and format to be incorporated in all application system level related erwin Data Modeler data models. As updates are applied to an Active Model Template, each application system data model (new from system design perspective, or, existing data model investments) can have an Active Model Template attached to manually or automatically update the application data model to the metadata defined in the Active Model template. From a data governance perspective, there is confidence that all this important metadata is consistent across your system metadata investments and supports communication and understanding among varying viewing audiences - both from business and IT technical perspectives.
Design Layer Architecture	Modeling Function	The Design Layer Architecture functionality in erwin Data Modeler supports separating and integrating the business or logical side of the data model and links these models together to support the synchronization of these data model artifacts. Although there are many different uses for this technique, this is most commonly used in situations where there is a need to include conceptual, logical and physical modeling techniques to separate the lines of responsibility during SDLC phases of application development and provide reuse capabilities among the various project phases. If this practice is leveraged in erwin Data Modeler, these model to model relationships are easily presented in the erwin Web Portal application as semantic relationships among these data models from a graphical as well as a textual view.
Source-To-Target Mapping	Modeling Function	The Source-To-Target Mapping facility in erwin Data Modeler provides ability to identify the mapping criteria needed to move data from one application system to another (i.e. Data Warehousing, MDM, etc., type initiatives). The target data model is populated with the source table and column objects that are imported from source data models or optionally from source system database catalog structures. The data modeler can then identify, track and manage the data source columns that are feeding an "Extract, Transform and Load (ETL)" process along with a business understanding of the "Transformation" rules. Using this feature can provide meaningful explanations from a business perspective for where, how and why the source column data values are being pulled from their source systems to feed the target data structure initiative. During any project design phase, the project team can document these mappings and report to project stakeholders to clarify the mapping requirements and provide deliverables to ETL personnel to implement what is agreed to. For completed system investments, this mapping can be retrofitted into the target data model to obtain a source of truth for the mapping process and also support enhancement and maintenance project efforts ongoing. This kind of effort can provide the mechanism to presenting graphical and textual impact analysis and data lineage at the column level in the erwin Web Portal very easily. The erwin Web Portal does provide the ability to apply this mapping directly, but this needs to be added post implementation. A best practice would be to leverage the Source-To-Target mapping in erwin Data Modeler to direct responsibility to the project team creating this effort. The data mapping facility available in the erwin Web Portal provides the capability to map data objects from different data modeling tools or where retrofitting existing mapping investments with erwin Data Modeler is not possible.

Import/Export Bridges	Modeling Function	There are many products in the marketplace that collect, store and reference metadata from a business perspective. ETL BI, UML products as well as the newer technologies like "Big Data", NoSQL, etc. that need to deliver this metadata into the hands of personnel that directly use these products in order to provide the business as well as the technical perspective for data values presented in these products. erwin Data Modeler provides over 120 import/export bridges that can synchronize both business and technical metadata among these products so that the same information is communicated among these products. By setting up erwin Data Modeler as the "Record of Source" for all metadata, this information can be easily replicated across all these vendor products to ensure everyone is looking at the same information regardless which vendor product is being used. From a data governance perspective, many of the various data formats that store metadata can be consolidated to a single medium - the erwin Data Modeler. When data models are integrated with other data governance practices as in the erwin Web Portal Data Governance edition, Data Governance Business Glossary analysts have all the metadata within the context of the supporting business application systems at their fingertips to derive accurate, clear, concise and complete corporate business terms.
Model Name and Description Subject Area Name and Description Diagram Name and Description	Modeling Object	erwin Data Modeler was the first product to support an integrated approach to defining data structures along with providing a business understanding for the various technical data structures supporting business applications systems. Since its inception during the early decades of relational technology advances, metadata related to the objects listed here have been present in the product. By providing good business names and definitions for these objects in the data model, this metadata is readily available in a data governance product like the erwin Web Portal.

Connect with us at **sandhill.co.uk**