



5 Data Modeling Tips for Better Data Governance

Karen Lopez, InfoAdvisors
April 2020



InfoAdvisors

2020 © InfoAdvisors - All Rights Reserved

Karen Lopez
@datachick

www.datamodel.com

#TEAMDATA®

IDERA Inc. sponsored this white paper.
ER/Studio is a registered trademark of Embarcadero Technologies,
Inc., a wholly-owned subsidiary of IDERA Inc.
TeamData is a registered trademark of InfoAdvisors.

Better Data Models, Better Data Governance

Are your data models truly supporting data governance?

Most data professionals have been involved in some way in data governance programs at some point in our careers. These programs range from light governance to fully staffed, well-funded C-level offices to ensure data quality and data protection for the enterprise.

No matter what level of data governance you have been involved with, data models provide a vital resource for successful governance programs. In this work, we look at the following data modeling features that need to be part of an enterprise's data modeling portfolio to ensure compliance with better data governance processes.

Data Governance Tasks Are Completed Faster with Better Data Models

In my experience, when teams are working toward the same goal, both individual and collaborative tasks are completed faster. Logical data models that have had strong engagement data stewards comprise hundreds of hours of decisions made by business users and data professionals. The results of these collaborative efforts can save a significant amount of data governance time.

Collaborative Tasks Are Easier

When there is less contention and more trust among team members, tasks are easier to complete because there are fewer distractions. Happier teams have better outcomes.

Data Model Quality is Data Quality

The quality of data models has a direct impact on data quality. Data models, as requirements and specifications for data, are the standards against which we measure data quality. Barebones data models, often just diagrams of databases, do not aid in better data governance.

The 5 Tips

Easy and Inexpensive Tips to Better Data Governance

Now that we have covered the benefits of better collaboration, let's look at how we can leverage good data modeling practices with good data governance. You may already practice a number of these, but ensuring you are doing all of them increases the value of your data models significantly and, therefore, your organization's value in you.

Let's first start with a definition of data governance from the *DAMA Data Management Body of Knowledge*.

“Data Governance (DG) is defined as the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets...While the driver of data management overall is to ensure an organization gets value out of its data, Data Governance focuses on how decisions are made about data and how people and processes are expected to behave in relation to data.”¹



Tip 1: Build Tailored Model Layouts for Data Stewards and Stakeholders

Data stewards are individuals who manage data governance in the field. They may be responsible for all the activities of data governance, or just one aspect, such as data quality, technical implementations of requirements, or data security.

Too many project teams, due to time constraints, prepare only one display of their data models. Often, this layout and markup are tailored for technical consumers of the models. But if data professionals focus models on technical audiences, these models form obstacles to data steward engagement.

Stewards: Hide non-business requirements

Technical objects and specifications can be obstacles to reviewing data models for business requirements. Some of the items that should be hidden for business reviews:

- Indexes
- PK and AK notations
- Surrogate keys
- Database owners and schemas
- Table names and properties
- Constraint names

¹ DAMA. Earley, S., & Henderson, D., Sebastian-Coleman, L (Eds.). *The DAMA Guide to the Data Management Body of Knowledge (DAM A-DM BOK)*. Bradley Beach, NJ: Technics Publications, LLC. 2017.

- Database Triggers
- Relationship triggers
- Technical columns (such as GUIDs, system dates)

In Figure 1 - Business Layout versus Everything Layout, note how much more focus a data steward has on business requirements between the two options for an entity.

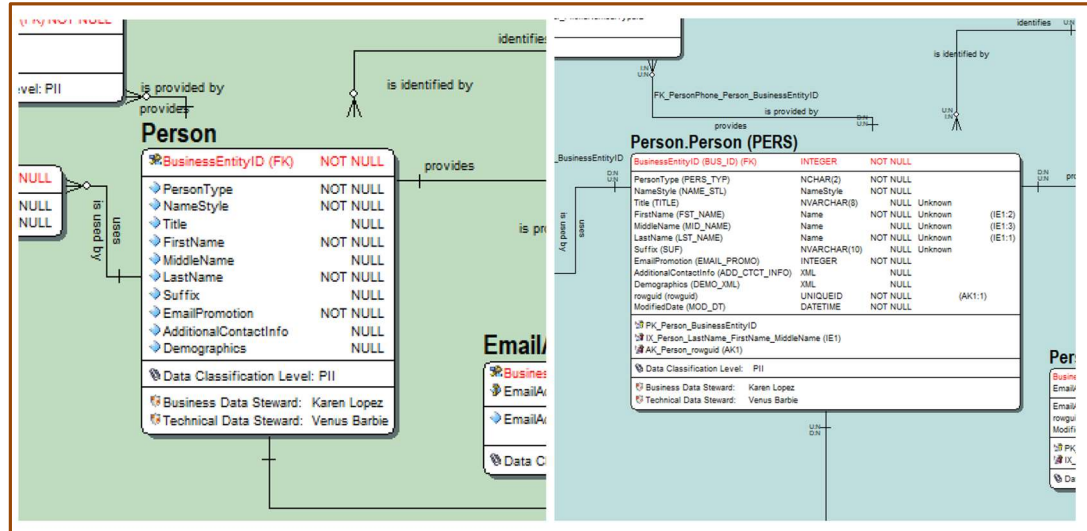


Figure 1 - Business Layout versus Everything Layout

There may be cases with certain data stewards, where adding more information may be warranted. For instance, if you use generic data types in your logical models, include them. If you need to review default values, add them. Incorporate just enough information to focus a review and approval results for better quality data models and faster results.

If time is a constraint to managing several views of a model, leverage data modeling tool features for templates, subject area inheritance, and macros to make preparing and managing multiple subject areas more efficient. The payoff for data steward success through the removal of distractions is significant.



Tip 2: Assign Data Privacy and Sensitivity Classifications

Consumers of data must be clear on usage requirements and constraints early on to avoid exposing the company to risk. A critical responsibility that a data modeler has is in classifying entities and attributes. *Privacy and sensitivity classification* involves assessing attributes for their sensitivity levels. As penalties from newer privacy and security legislation arise, the need to classify and model these levels becomes more significant.

The last tip discusses an additional type of data classification that of grouping data by business terms.

Use Data Privacy and Sensitivity Classification Tools to Assist

Traditionally we data professionals have inspected each data attribute to determine which elements are Personally Identifiable Information (PII), financial data, health sensitive data, or confidential data. One must understand the meaning of the data to know whether it is sensitive data. Who better to understand data than the data architect and a business steward?

Now we have tools that can assist us in identifying the classification of data attributes.

Most RDBMS vendors include data classification features within their database client tools. These classifications are usually accomplished by examining database column names to infer a sensitivity type. Figure 2 - *SQL Server Data Classification Recommendations* shows an example of how Microsoft SQL Server Management Studio (SSMS) makes data sensitivity recommendations based on column names in a database. These types of tools are highly dependent on having meaningful column names. Data classifications are reviewed and stored in the database as metadata. This metadata should be compared into physical and logical data models and vice versa.

The screenshot shows the 'Data Classification - GalacticWorks' window in SQL Server Enterprise Manager. It displays a table named 'Employee' with columns: 'DOB', 'AddressLine1', 'AddressLine2', 'AddressLine4', 'City', 'PostalCode', 'EmailAddress', 'PasswordHash', 'PasswordSalt', and 'FirstName'. Each column has a recommended 'Information Type' and 'Sensitivity Label'.

Schema	Table	Column	Information Type	Sensitivity Label
HumanResources	Employee	DOB	Date Of Birth	Confidential - GDPR
Person	Address	AddressLine1	Contact Info	Confidential
Person	Address	AddressLine2	Contact Info	Confidential
Person	Address	AddressLine4	Contact Info	Confidential
Person	Address	City	Contact Info	Confidential
Person	Address	PostalCode	Contact Info	Confidential
Person	EmailAddress	EmailAddress	Contact Info	Confidential
Person	Password	PasswordHash	Credentials	Confidential
Person	Password	PasswordSalt	Credentials	Confidential
Person	Person	FirstName	Name	Confidential - GDPR

Figure 2 - SQL Server Data Classification Recommendations

Other data sensitivity tools can perform data profiling (examining data itself) to make data classification recommendations. Both types of services are essential to efficiently classifying data. However, data stewards, together with data modelers, still need to categorize data by inspecting each attribute. For instance, a special meal request or a wheelchair request by a traveler may even be considered health sensitive data.

Once data classifications have been completed, data models can be updated with this critical information.

3

Tip 3: Extend and Enhance Data Entities with Metadata

Many data models begin their lives as a reverse-engineering of a database. These models contain only those objects and properties a database contains, namely data structures and constraints. Yet useful data models, especially logical data models, are enhanced by business-related metadata:

Use extended definitions and notes

Time-constrained data professionals often find it difficult to develop meaningful data modeling object definitions, but the payback for doing so is significant over time. The best time to write a meaningful and complete definition is at the time the object is created in the model. To make entity definition writing faster, consider using the format of:

A {noun} that {verb phrase + context}. This includes {more detail} and excludes {more detail}.

Data professionals that follow a pattern in writing definitions find that they are faster to write and better understood. One of the biggest myths in writing definitions is that one does not need them if they have good naming standards.

Refrain from listing attributes in the definition, as attributes may end up in other entities or be renamed later. It is also recommended to exclude references to other entities, except in referring to their business concepts: *invoices*, not *INVOICE*, as an example.

Developing meaningful and extended definitions that appear in the models, portals, data catalogs, and even database objects is a tremendous value in supporting data governance throughout the development process. Idera's ER/Studio Data Architect supports hover-based tips that show definitions. An example of this presentation is shown in *Figure 3 - Entity Definition Display*.

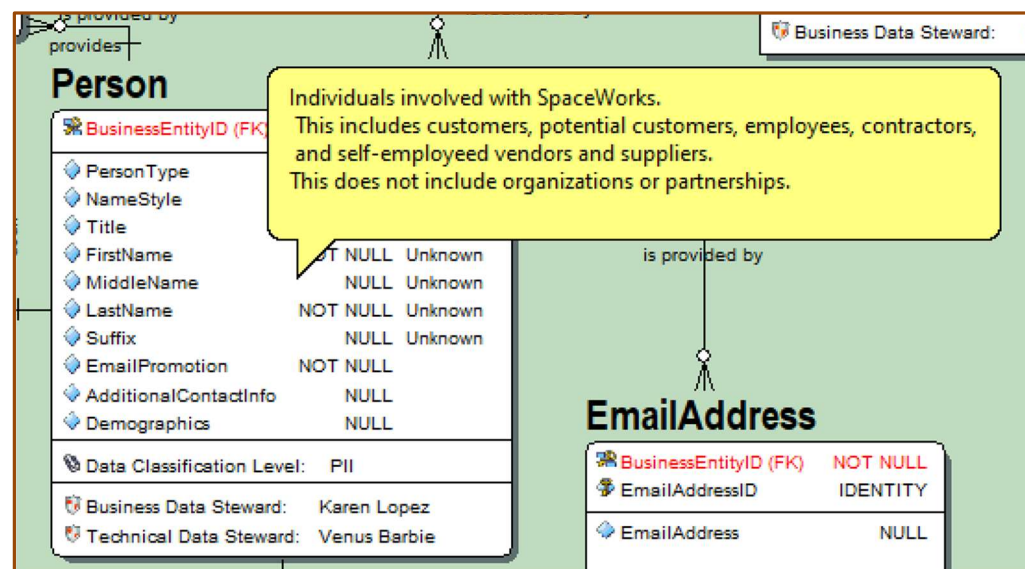


Figure 3 - Entity Definition Display

Include Data Steward Information

Since data stewards are responsible for managing data governance processes and policies, ensure that modeling objects contain stewardship information. In some organizations, there is just one stewardship role, but in more mature data governance programs, there may be many. These might be divided into strategic, tactical, and operational positions or in business and technical roles. No matter which formula your data governance areas follow, including the metadata ensures that everyone with a question or concern knows whom to contact. See *Figure 4 - Stewardship Metadata* for an example of Business Data Steward and Technical Data Steward metadata in a Logical Data Model.

Including this metadata in models also means it can carry forward to data portals, data catalogs, and data itself.

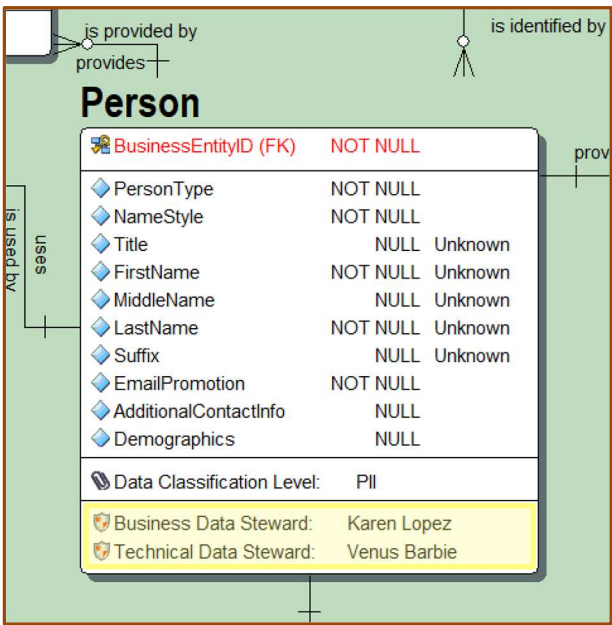


Figure 4 - Stewardship Metadata

Include Data Privacy and Sensitivity Classifications

All the work done in Tip 2, Assign Data Classifications, is an absolute must-do in data modeling. Data privacy and sensitivity classification metadata is an essential aspect of any review or discussion about data. From C-level reviews to developer implementation of applications, data classification should be at the forefront of each stakeholder's thinking.

In *Figure 5 - Data Privacy Classification*, see how data privacy metadata is shown in a data model

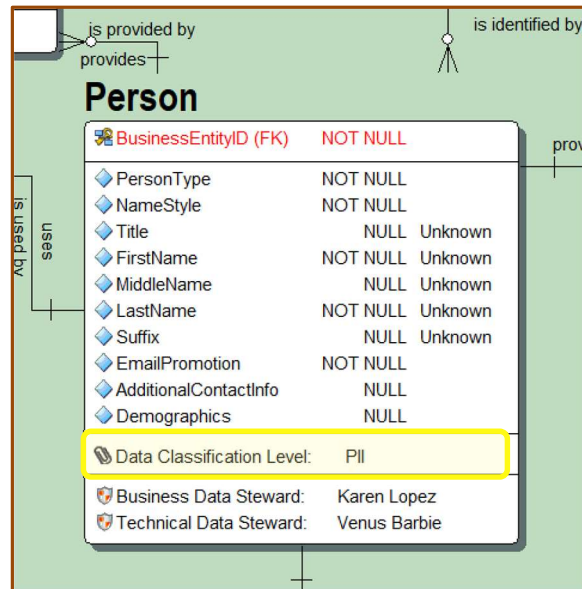


Figure 5 - Data Privacy Classification

Use Other Visual Artifacts

One of the best ways to leverage data modeling tools is to make use of text and shapes to draw attention to specific concepts or to enhance the review process. Sometimes data professionals do all this annotation in other tools using screen captures of data models. However, leaving the annotations in other tools means they are lost to the data model and usually have to be recreated for future discussions. Since these are independent objects in a model, they can be hidden or included as needed. Including them in a model also means they are subject to version control and backups. *Figure 6 - Entity Note* shows an example of a text note added to an entity to clarify why it is used in the model.

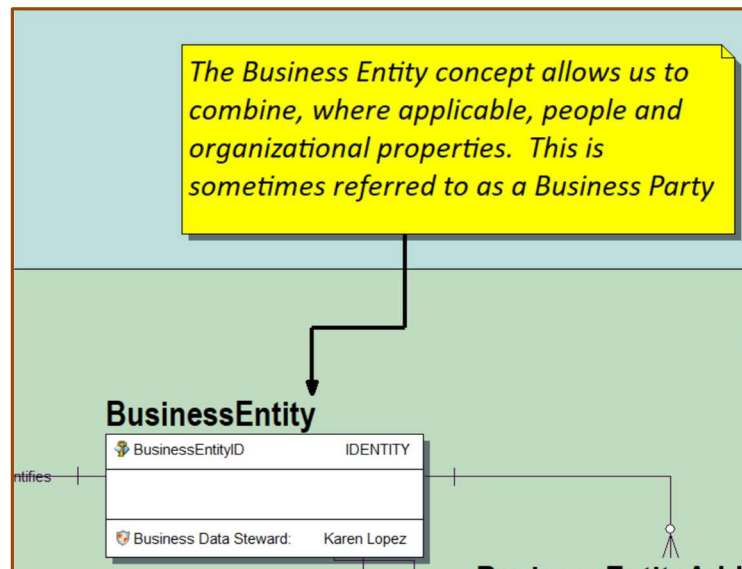


Figure 6 - Entity Note

Use Business Data Objects to Group Entities

One of the mismatches between how data stewards think of data and how data modelers do is one of data normalization. To a data steward, an *Invoice* is a document that contains all the data on a paper or electronic record of a demand for payment. To a data professional, *Invoice* is just one entity of a collection of entities that comprise that document. These points of view, while different, are easier to manage with the use of Business Data Objects. These are non-technical groupings of entities in a logical data model. Figure 7 - *Person Business Data Object Expanded* shows how Person entities are contained in a business object named Person Object. This object can be collapsed as in Figure 8 - *Person Business Object Data Collapsed* to hide complexity when needed and expanded as in the figure to include the details of the entities.

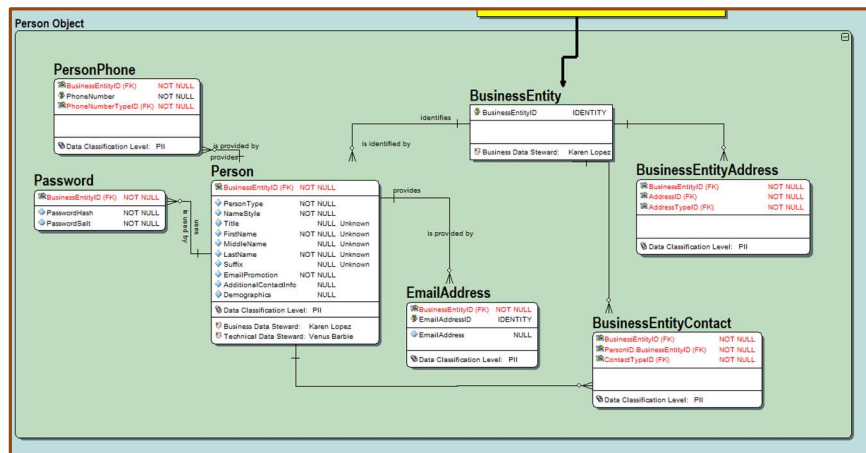


Figure 7 - Person Business Data Object Expanded

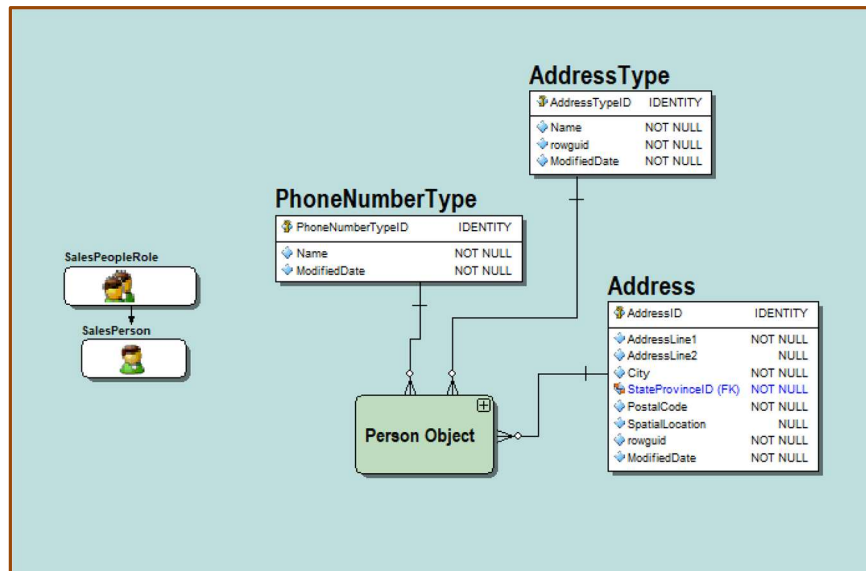


Figure 8 - Person Business Object Data Collapsed

**Tip 4: Model Data Security Requirements**

Gone are the days that we data modelers put together a logical data model, generated a physical model based on it, then generated a database script and threw it over to a database administrator (DBA) to worry about how to protect that data.

Data protection requires data classifications, compliance, and business reviews to identify which data needs extra security. Then we need to model how to secure that data. All these things must be documented across conceptual, logical, and physical data models. This attention to security ensures that the models can adequately support model-driven development that is not dependent upon the memory of a developer or DBA to apply the security after the fact.

Add Business Security Requirements to Logical Data Models

Security is a data steward responsibility, which means it's a data modeler's responsibility. Data modelers may not need to specify how data is secured, but we do need to model security requirements:

- Data encryption requirements
- Data masking requirements
- Data access requirements, including attribute level and instance (row-level)

Data sensitivity classifications are where we should start modeling requirements. The business sets these requirements, often through the data governance program. With the assistance of data security teams, we might find that we must encrypt or mask specific columns. We might also need to specify which business roles should have the right to see unencrypted or masked data.

Add Technical Security Requirements to Physical Data Models

When working with DBAs, a data modeler needs to ensure that the security requirements are implemented thoroughly. The data modeler represents the data steward's need for good governance, and the DBA serves the operational needs to implement requirements that perform well.

Modeling data security designs can also contribute to ensuring ad hoc uses of data, such as self-serve Business Intelligence (BI) tools, carry the security requirements to business users of the data. A data scientist using customer data should be able to tell how the data she wants to use may be used, how it may have been masked, and how it should be protected.

**Tip 5: Implement Easy Data Model Collaboration**

The value of a data model comes from its regular use, not just by data modelers. Properly used data models can be easily leveraged by data stewards, business users, developers, DBAs, data scientists...the list could go on to almost everyone in an organization.

One single data model display does not meet all those needs. We have seen in earlier tips on how data models can be tailored to meet the needs of target audiences. But there are other ways effective modeling can meet different needs.

Include Comments and Questions

In the past, **model comments and questions** were often managed on paper notepads, e-mails, and spreadsheets. These comments were difficult to tie back to a specific object. Collecting and managing them required Herculean efforts, often by both the data steward and the data professional. We can now manage these in our data models. That means they can be versioned, shared, secured, and backed up. They **are attached to an object and can be seen by others who often have the same questions or comments**. Follow up is also more efficient as comments are tracked back to a specific account.

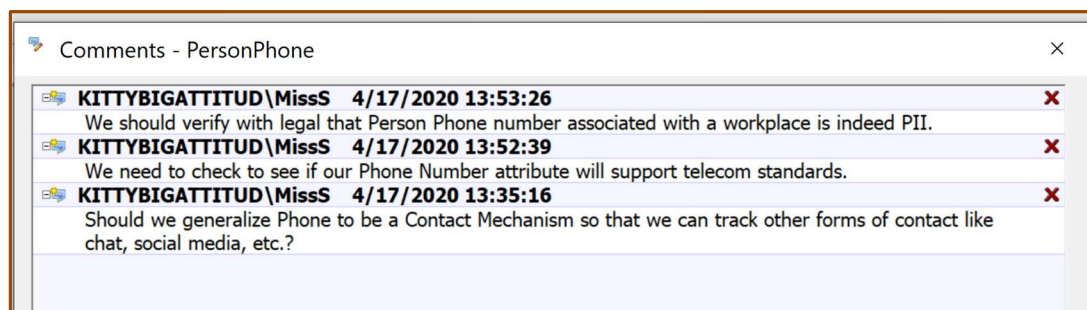


Figure 9 - Data Model Comments

Leverage Data Model Repository

The **ER/Studio Repository** provides version control at the property level of the data model object. It offers versioning, snapshots (named releases), and checking of data models. The check-in and check-out functions support sharing and collaborating on data modeling efforts at the lowest level of granularity required by data professionals. Data stewards, data modelers, DBAs, developers, and security analysts can work together on the same modeling objects, even at the same time without fear of overwriting each other's changes. As you can see in Figure 10 - Collaboration with Repository, data model work can happen at the same time, across many groups, and even remotely.

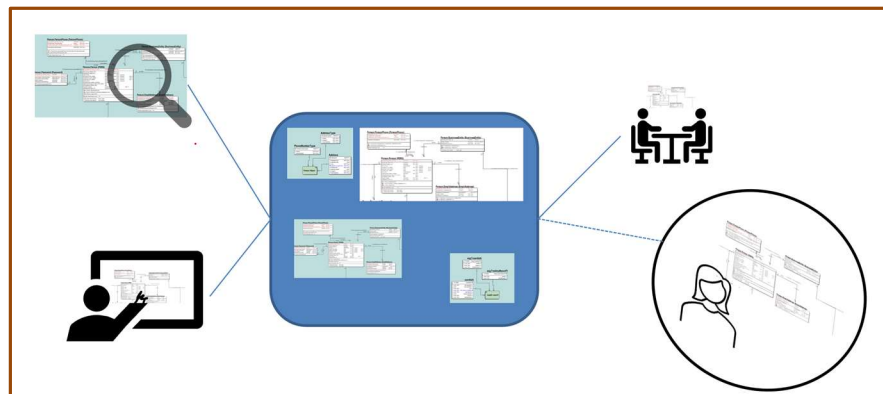


Figure 10 - Collaboration with Repository

BETTER DATA GOVERNANCE

Manually sharing and coordinating data modeling work at the file level is not possible on real-world-sized projects – there’s just too much complexity and reuse of objects. File-level version control is not fine-grained enough to support responsive data modeling efforts.

The Repository also includes security features to protect the integrity and security of data models. Data governance requires data model quality as well as data quality; having the right collaboration and protection tool is imperative.

Promote Models as the Go-To Record for Data Knowledge

Anyone working with data assets should use data models as the authority for business requirements about data. A data model provides knowledge on not only the structure of data assets, but also the meaning, rules, and guidance for that information. In support of these authoritative records, Idera Team Server also notifies users of changes to ‘followed’ data assets. If developers are concerned with a table in a database, they can follow it and be notified of any changes to that table are made.

Leverage Data Portals and Catalogs

Data Stewards and Data Modelers generally use different modeling paradigms to manage this knowledge. Data Stewards are more focused on Glossaries of Business Terms. Data Modelers concentrate on logical and physical models. Idera Team Server contains all these paradigms in the same repository and joins them together. Both roles need to be able to review and comment on all of those. One can define business glossaries with terms and relationships between them, then link those glossary terms to entities and tables in data models. This ability to drill down or up from glossary to implementation objects is critical to managing the effectiveness of data modeling initiatives. In Figure 11 - Business Glossary Terms, the list of terms and their metadata, including definitions, can be seen.

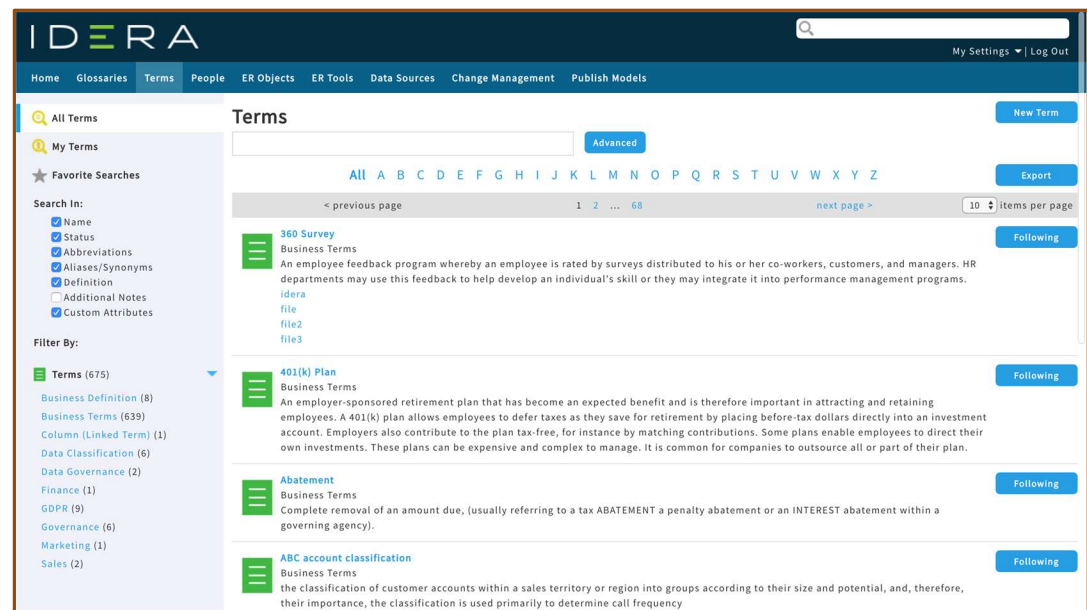


Figure 11 - Business Glossary Terms

One can drill down from a term to see the objects that are classified or characterized by that term. This traceability is the other type of data classification discussed in the first tip. In *Figure 12 - Term Found In*, the business term and its grouping of tables can be seen.

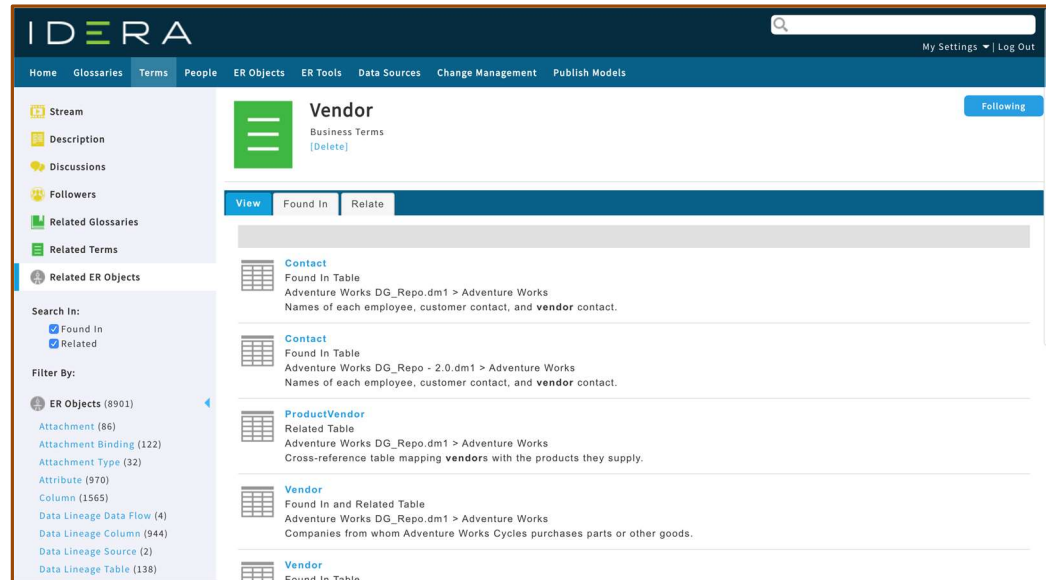


Figure 12 - Term Found In

This classification of data is a crucial part of data governance as it can be used for impact analyses and measuring data and data model quality. Furthermore, the roles involved in data now have a single continuous ecosystem to manage their data. The knowledge and requirements of Data Stewards can be connected and compared with that of Data Modelers and DBAs to maintain consistency throughout the process of managing data.

Finally...

While the tips here cover a range of topics from collaboration to technical implementations, they all share a common trait: the maturity of data governance in an organization impacts their ability to succeed. Good data modeling also affects how and where data governance work is completed. Collaboration, usability, completeness all form vital components of data governance efficiency. Within this harmonious ecosystem, as data stewards define standards around data, data modelers can ensure that data assets deliver against those standards.

New to Data Governance?

An organization starting its path towards formal data governance might be enhancing its logical data models with necessary metadata around data stewards and sensitivity levels.

This organization might also be starting a business glossary and implementing a data model

and standards portal. New data governance initiatives can leverage such knowledge from mature data modeling assets.

More Mature Data Governance?

A more mature organization with a formal, enterprise-wide data governance program might be managing these items in other systems and publishing the results of that work to their data models and portals. They might also be documenting new data governance items in their models, then publishing those out to their formal data governance tools. Where the work is done is less important than the fact that the activities are happening and being recorded then shared.

Being part of TeamData® means leveraging your data models to support data governance. No matter where your organization fits in these maturity models, your data models play a crucial role in ensuring your data governance is timely, engaging, and successful.

About the Sponsor



IDERA understands that IT doesn't run on the network – it runs on the data and databases that power your business. That's why we design our products with the database as the nucleus of your IT universe.

Our database lifecycle management solutions allow database and IT professionals to design, monitor, and manage data systems with complete confidence, whether in the cloud or on-premises.

ER/Studio is the collaborative data modeling solution for data professionals to map and manage data and metadata for multiple platforms in a business-driven enterprise data architecture.

Whatever your need, IDERA has a solution.

About the Author

Karen López is a Senior Project Manager and Architect at InfoAdvisors. She has more than twenty years of experience in helping organizations implement large, multi-project programs.

She tweets as @datachick.

InfoAdvisors is a Toronto-based project and data management consulting firm. We specialize in the practical application of project methods and tools. Our philosophy is based on assessing the cost, benefit, and risk of any technique to meet the specific needs of our client organizations.

InfoAdvisors offers data modeling training, including training for non-modelers such as business users, database administrators, project managers, and developers.

We would like to help you add value to your data models on #TeamData

For more information, visit www.datamodel.com.