By Steve Hoberman

Steve Hoberman & Associates, LLC

Leveraging the Industry Logical Data Model as Your Enterprise Data Model

Sponsored by:

TERADATA
Raising Intelligence

Table of Contents

Introduction 3

Section One: Enterprise Data Model Overview 4
- EDM Definition 5
- EDM Settings 5
  - Detail 6
  - Abstraction 9
  - Time 10
  - Function 11

Section Two: Use of an Industry Model 12
- Factors in Deciding Whether to Use an Industry Model 12
- Industry Model Resources 14
  - Websites 14
  - Books 14
  - Industry Organizations 15
  - Vendor Models 15

Section Three: Teradata Industry Logical Data Model Overview 16
- Communications Logical Data Model (CLDM) 19
- Financial Services Logical Data Model (FS-LDM) 19
- Healthcare Logical Data Model (HCLDM) 19
- Manufacturing Logical Data Model (MLDM) 20
- Media & Entertainment Logical Data Model (MELDM) 20
- Retail Logical Data Model (RLDM) 20
- Transportation and Logistics Logical Data Model (Trans LDM) 20
- Travel and Hospitality Logical Data Model (THLDM) 20
- Utilities Logical Data Model (ULDM) 20

Section Four: iLDM Evolution through Unification 21
- Advantages of Unification 21
- Foundational and Functional Unification Building Blocks 22

Section Five: Areas that Benefit from the EDM and iLDM 23
- Enterprise Resource Planning 24
- Master Data Management 24
- Business Intelligence 25
- Data Quality 26
- Employee Development 26

Section Six: Use of the iLDM 27
- Approach 27
  - Identify EDM Objectives and Settings 27
  - Produce the Business Solution (the Subject and Logical Data Model) 27
  - Produce the Technical Solution (the Physical Data Model) 29
  - Maintain EDM 29
- Quicker ROI with an iLDM 29

Section Seven: EDM Advice 31

About the Author 32

References 32

Introduction

"Data is one of an organization's most valuable assets. All current and future business decisions hinge on data. An EDM is essential for the management of an organization's data resource. The core principle of data management is order; applying order to the vast universe of data. To manage data is to apply order. According to the second law of thermodynamics; the universe and everything in it, continually heads toward chaos; it takes energy to bring order. The same holds true for data, left alone, it continually deteriorates to a state of disorder. It takes concerted effort to keep data in order. An EDM brings order."

Noreen Kendle, Enterprise Data Architect, Delta Airlines

Every profession has a pinnacle of success, that if somehow achieved elevates you into the elite and enviable few. An astronomer discovers a far off galaxy. A biochemist invents a vaccine. An archeologist uncovers a lost city. A CIO achieves return on investment (ROI) on an enterprise data model (EDM).

The concept of an EDM is more than 20 years old. Yet during those 20 years, there have been very few success stories. Very few organizations have demonstrated a ROI within a reasonable amount of time. I believe, in many cases, the reason behind a low success rate is that many organizations lack group agreement on the costs and benefits that are needed to calculate ROI. EDM cost and benefit need to be agreed upon before the EDM work starts.

Time and money requirements for the EDM need to be budgeted for and monitored during the EDM's construction and maintenance. Throughout this white paper, I will refer back to a survey performed on 22 organizations that have built and are using EDMs. A common theme across responses is that very few organizations know the cost of an EDM. Only four out of 22 organizations could quantify costs. Actual responses included "I don't know," "blank check," and "I did not keep track of time." As with any application development effort, we need to budget and monitor costs both in terms of money and time.

To calculate ROI you also need to know the benefits. Very few organizations know the concrete benefit of an EDM prior to starting the effort. Consider a conversation I had with an IT senior manager a number of years ago at a Wall Street brokerage firm. We were talking about our data chaos, lacking standard definitions for our data elements, and not having a complete list of applications where each data element resides. The senior manager shared with me his dream of covering one entire wall in his office with a list of all data elements in our organization and the systems where each data element is used. However, what is the benefit behind this single view of the information? There is a difference between knowing the deliverable and knowing the benefit of the deliverable. We need to know the business benefit before starting an EDM effort.

The next step after planning the upfront costs and benefits for your EDM is to decide what EDM settings are appropriate and on the usage of an industry data model. There are a number of EDM settings and a number of ways of leveraging industry data models. Following the maxim that "knowledge is power," you will need to become aware of the possibilities and then map what is most appropriate to your proposed costs and benefits. This white paper will explain the EDM and its settings, and the different ways organizations can leverage industry data models, specifically the Teradata® Industry Logical Data Models (iLDMs). The goal of this paper is to increase your awareness of what the EDM and iLDMs provide, and therefore increase your chances for success and for joining the enviable few who have demonstrated EDM ROI.

Section One provides an overview to the EDM. We will define the EDM and then discuss four settings that lead to many EDM variations. Section Two covers the factors in deciding whether to build an EDM from scratch or use an industry data model. Industry data model resources are discussed. Section Three provides an overview to the iLDMs in general and briefly explains each of the nine iLDMs: Communications, Financial Services, Healthcare, Media & Entertainment, Manufacturing, Retail, Transportation and Logistics, Travel and Hospitality, and Utilities. Section Four explores Teradata iLDM evolution through unification. Section Five explains several initiatives that require an EDM for success, including Enterprise Resource Planning (ERP), master data management, business intelligence, data quality, and employee development. Section Six talks about ways in which organizations are leveraging the Teradata iLDMs, including approach and results. I'll conclude this paper with a "Top 10" list of advice about the EDM.

Section One: Enterprise Data Model Overview

A model is a set of symbols and text used to make a complex concept easier to grasp. The world around us is full of obstacles that can overwhelm our senses and make it very challenging to focus only on relevant information needed to make intelligent decisions. A map helps a visitor navigate a city. An organization chart helps an employee understand reporting relationships. A blueprint helps an architect communicate building plans. The map, organization chart, and blueprint are all types of models that represent a filtered, simplified view of something complex.

A data model uses symbols and text to help developers and analysts understand a set of data elements and their business rules better. An increase in understanding the data leads to an improved communication level within the organization and therefore a more flexible and stable application environment. The symbols on a data model take the form of boxes and lines. The boxes are called "entities" and represent concepts of importance to the business, such as Customer, Product, and Order. The lines are called "relationships" and represent the rules that constrain the entities.

Every model has a defined scope. A map might be limited to New York City, an organization chart to the IT department, and a data model to a specific functional area such as order processing. A model can also have a very broad scope and represent *everything.* A map can represent the planet, an organization chart the corporation, and a data model all of the information within an entire organization. A data model that captures all of the information within an entire organization is called an enterprise data model (EDM).

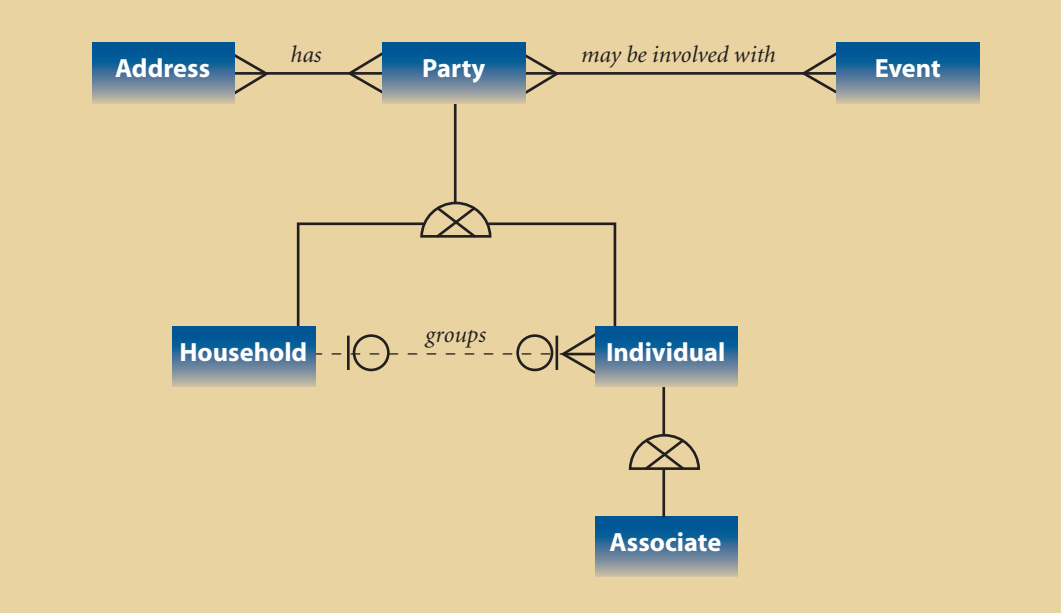

Figure 1. Subset of Teradata Manufacturing Industry Conceptual Data Model.

EDM Definition

An EDM is a subject-oriented and integrated data model containing all of the data produced and consumed across an entire organization. Subject-oriented means that the concepts on a data model fit together as the CEO sees the company, as opposed to how individual functional or department heads view the company. There is one Customer entity, one Order entity, etc. Integration goes hand in hand with subject-orientation. Integration means that all of the data and rules in an organization are depicted once and fit together seamlessly. Every data element has a single definition and name. Integration implies that with this single version of the truth comes a mapping back to the chaotic real world. For example, if Customer Last Name lives in 10 applications within an organization, the integrated view would show Customer Last Name only once, and in addition, capture the mapping back to these 10 applications.

An EDM, sometimes called a "corporate model" or "business model," comes in many shapes and sizes, as we'll see in the next section. To give you a taste for an EDM and review the meaning of the data modeling symbols, see Figure 1. This is part of Teradata's Manufacturing Industry Conceptual Data Model (CDM).

There are six entities in this model: Address, Party, Event, Household, Individual, and Associate. The lines on this model represent relationships that enforce business rules. You only need to know a handful of symbols to read a data model. On a relationship line, the small vertical line means "one," the circle means "zero," and the triangle with a line through the middle means "many." Some people call the "many" symbol a crow's foot. The text labels on the relationships help in interpreting the business rule. The semi-circle with an "x" inside represents a subtype symbol. Subtyping is when the same data elements and relationships across two or more entities are grouped together, instead of repeated for each entity. For example, Party contains all of the common data elements and relationships for Household and Individual. Party is called the supertype. Household and Individual are the subtypes. Subtyping is a common technique on EDMs.

So here are the relationships captured on this EDM subset:

- A Party has many Addresses.
- An Address belongs to many Parties.
- A Party may be involved with many Events.
- An Event is involved with many Parties.
- A Party can be either a Household or Individual.
- An Individual can be an Associate.
- A Household groups one or more Individuals.
- An Individual can be grouped into a Household.

EDM Settings

A camera has many settings available to take the perfect picture. Imagine facing an awesome sunset with your camera. With the same exact sunset, you can capture a very different image based on the camera's settings, such as the focus, timer, and zoom. You might, for example, zoom out to capture as much of the sunset as possible, or zoom in and focus on people walking by with the sunset as a backdrop. It depends on the purpose of the photograph.

There are four settings on a camera that translate directly over to the EDM: zoom, focus, timer, and filter. See Figure 2 for these settings and to view how many organizations in our EDM survey have this setting for their EDM. An EDM is characterized by one setting value from each setting. The bar chart next to each setting indicates the number of organizations in my survey that have this particular setting value. Which setting is right for your EDM? As with photographing the sunset, it depends on the purpose. Match the goals of your EDM with the appropriate EDM settings.

The zoom setting on the camera allows the photographer to capture a broad area with minimal detail, or a detailed area without much context. Similarly,

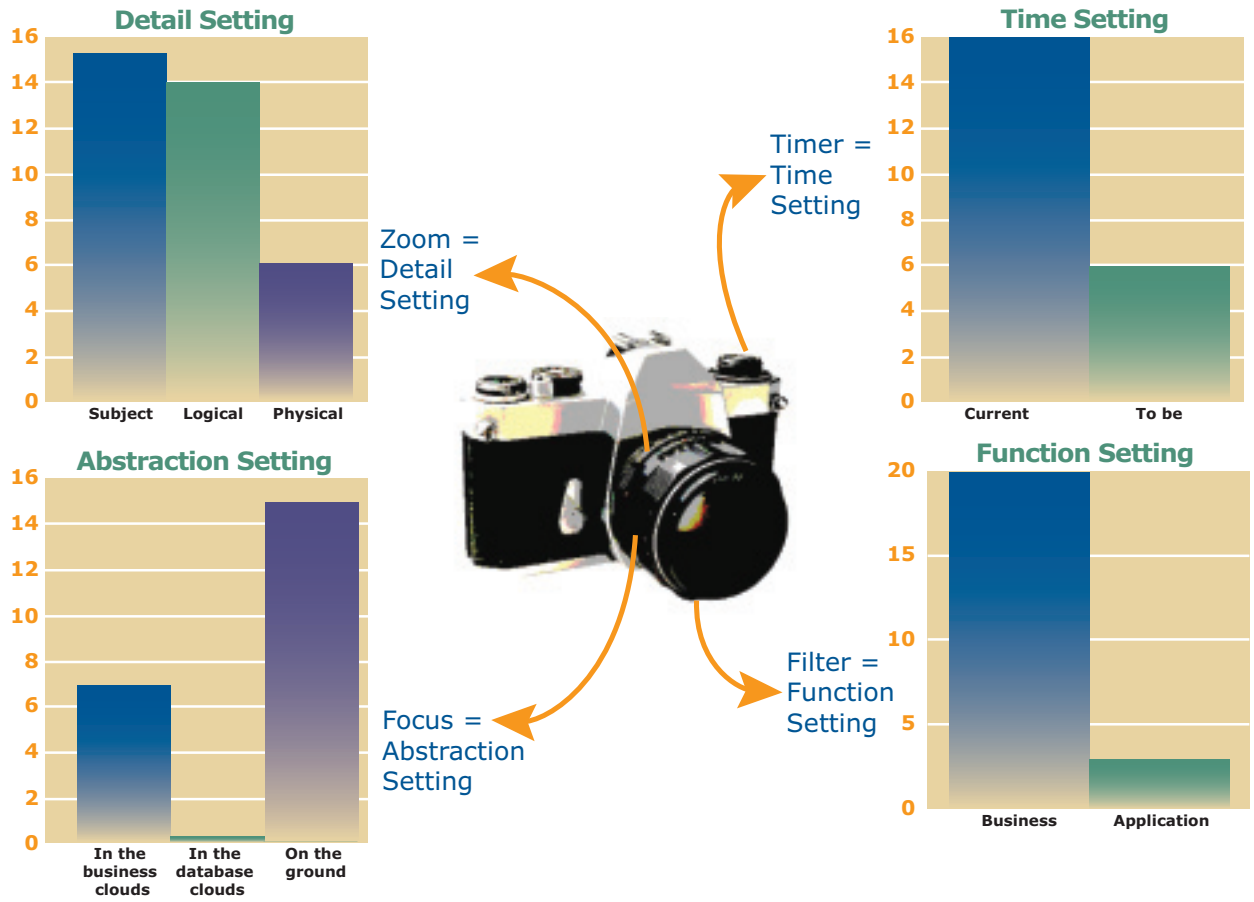

Figure 2. EDM settings and survey results.

the detail setting for the EDM varies how much you see in the picture, as well. A majority of the organizations in our EDM survey have an EDM with the widest zoom – the subject level. The focus setting on the camera can make certain objects appear sharp or blurry. Similarly, the abstraction setting for the EDM can use generic concepts, such as Party and Event to blur the distinction between concepts. The most common response in our survey was for the sharpest focus, the on-the-ground view. The timer allows for a real-time snapshot or a snapshot of some time in the future. Similarly, the time setting for the EDM can capture a current view or a view some time in the future. The most common response in our survey was for the current view. The filter setting can taint the appearance of the entire picture to produce a certain effect. Similarly, the function setting for the EDM taints the EDM picture with either a business or application view. The most common response in our survey was for the business view.

Detail

Similar to how the zoom on a camera allows you to focus on a small part of a scene or pan out to capture all of the details, the detail setting for an EDM allows you to represent some or all of the details on an EDM. The detail setting adjusts the level of detail for an EDM, making the EDM either at the very broad and

Table 1. Detail setting values.

Setting value	Choose this setting value for your EDM when you…
Subject	**Need a starting point.** This level works well when you see that there are many integration issues and it's very difficult to get people to agree on what terms mean or where data should be sourced from. For example, this level was ideal for a university I worked with, where we couldn't gain agreement in the room as to the definition of a student. **Have very little time.** A subject model takes a relatively short amount of time to build. The modeler just needs to include basic and critical concepts within the scope, and doesn't need to include data elements or think about normalization or database constraints. In addition, because it's at a high level, it also takes a minimal amount of effort to keep up to date. **Need a high-level understanding.** This level is sufficient if all you need is to get people from different areas to agree on what terms mean or to resolve high-level sourcing issues.
Logical	**Need a business solution.** This level should be chosen if it's important to get down to the details of identifying all of the data elements, and you really need to understand how something works independent of technology. **Need a prerequisite to the physical design.** If you eventually have to get to database tables, a logical level should be chosen before the physical. **Reengineer an enterprise.** If you plan to replace one or more legacy systems or introduce a large new initiative, such as business intelligence (BI), choose this level.
Physical	**Need a technical solution.** If you need to create a database for an application. This is the final end product after all of the logical work is complete and physical techniques, such as denormalization and indexing, have been applied. For an EDM, a physical setting works well if you plan to have a single application drive your business. **Reverse engineer the business.** As a first step towards replacing applications, all of the existing applications should be modeled.

high-level subject view, or the more detailed logical or physical views. Some organizations have a combination of subject, logical, and physical EDMs, depending on EDM purpose.

Subject. A subject model represents the enterprise at a very high level. It's a very broad view containing only the basic and critical concepts for a given scope. Here, *basic* means that it's usually mentioned one hundred times per day in normal conversation. *Critical* means that without this subject, the department, company, or industry would be greatly changed. Some subjects are common to all organizations, such as Customer, Product, and Employee. Other subjects are very industry- or department-specific, such as Telephone Number for the telecommunications industry or Trade for the

brokerage industry. I built a subject EDM for a manufacturer that represented their entire business in about 150 subjects. I built a subject EDM for a university that represented their entire organization in about 100 subjects. The model in Figure 1 is at the subject level.

Logical. A logical data model represents a detailed enterprise *business* solution. In our relational world, the business solution takes the form of a normalized model. Normalization is the process of ensuring every data element on a data model is single-valued and depends completely and only on the primary key. Data elements are grouped into sets according to how they are uniquely identified. On a normalized model, the business rules governing every data element need to be fully captured and validated. For example, a data element called Student Name on a normalized model will most likely be parsed into Student First Name and Student Last Name and be grouped with similar data elements into a Student entity with a primary key of Student Identifier. A logical EDM is independent of application technology, such as database, network, or reporting tools. A logical EDM can be a very large model, with thousands of entities, relationships, and attributes. I worked on a logical EDM for a telecommunications company that contained more than 3,000 entities! It's important to distinguish a logical from a semantic data model. Whereas an LDM represents an *enterprise* business solution, a semantic data model represents an *application* business solution. The single enterprise LDM can be mapped to multiple semantic data models, each representing different application requirements' slices of the EDM. Student, for example, will be modeled fully normalized on the enterprise LDM, yet can appear in a dimensional structure on a semantic data model for a data mart application.

Physical. A physical data model represents a detailed enterprise *technology* solution. It's optimized for a specific context, such as software, hardware, or data. A physical data model resembles its LDM, the exception being the addition of performance techniques such as denormalization, indexing, views, and partitioning. The same LDM can look completely different depending on how it has been optimized. The same order LDM for example, will have completely different order processing and sales reporting physical data models. Order processing involves lots of inserts and updates, and sales reporting involves lots of retrieval, and therefore, different physical data modeling techniques will be used. A physical EDM is rare because this would mean that all of these conflicting usage patterns will need to be documented. The easiest way to obtain a physical EDM is if an organization has a single application. This might sound farfetched, but I've worked with an entertainment company that had all of its data stored in SAP/R3, which made the physical data model much more attainable.

As mentioned in the introduction to this paper, ROI becomes easier to achieve when you know the purpose of your EDM and match it appropriately to the right setting value. Table 1 shows each setting value for Detail, along with when to choose each setting value. Think about the purpose of your EDM, and then select the right level or levels. Before making your selection with each of these settings, you also might consider reading Section Five, which describes the five initiatives that benefit from the EDM.

Table 2. Abstraction setting values.

Setting value	Choose this setting value for your EDM when you…
In the business clouds	**Need to resolve broad business issues.** If you're finding it extremely difficult to get people from different departments to agree on definitions and terminology, using very generic concepts, such as Person and Organization, can lead to across-the-board agreement. **Need a starting point.** Before embarking on more detailed discussions, you can get people to agree on the basics first. **When you need something in a hurry, and it has to be right!** If you aren't sure what the requirements are due to time or resource constraints, using generic concepts can guarantee you have the right answer. For example, if you don't know all of the different types of organizations, the concept of Organization can embrace them all.
In the database clouds	**Need to resolve broad technical issues.** If you're finding it extremely difficult to get people from within IT to come together with one set of terms, using database concepts, such as Object and Attribute, will lead to across-the-board agreement. **When you need something in a hurry, and it has to be right!** If you aren't sure what the requirements are due to time or resource constraints, using generic concepts can guarantee you have the right answer. Just ask anyone who has ever designed a metadata repository.
On the ground	**Need detailed business understanding.** If you need some abstract concepts for integration yet the majority need to be concrete terms to really get people to understand, then this is the correct level.

Abstraction

Similar to how the focus on a camera allows you to make the picture sharp or fuzzy, the abstraction setting for an EDM allows you to represent sharp (concrete) or fuzzy (generic) concepts. Abstraction brings flexibility to your data models by redefining and combining some of the data elements, entities, and relationships within the model into more generic terms. Abstraction is the removal of details in such a way as to broaden applicability to a wide class of situations while preserving the important properties and essential nature from concepts or subjects. By removing these details, we remove differences, and therefore, change the way we view these concepts or subjects, including seeing similarities that were not apparent or even existent before. Abstraction can exist on an EDM as "in the business clouds," "in the database clouds," or "on the ground."

In the business clouds. At this level of abstraction, only generic *business* terms are used across the EDM. The business clouds model hides much of the real complexity within generic concepts such as Person, Transaction, and Document. In fact, both a candy company and insurance company can look very similar to each other using business cloud entities. If you lack business understanding or don't have access to business documentation and resources, an EDM "in the business clouds" can work well.

Table 3. Time setting values.	
Setting value	**Choose this setting value for your EDM when you…**
Today	**Need a starting point.** Before understanding what needs to be changed, you should first understand the current situation.
Tomorrow	**Have an application milestone in near future.** If sometime in the not-so-distant future, the organization is planning a major software rollout or upgrade. **Need to perform impact analysis.** If you need to show how something large, such as a major initiative or reorganization, will impact the organization, a *tomorrow* view overlaying a *today* view is preferred.

In the database clouds. At this level of abstraction, only generic *database* terms are used across the EDM. The IT clouds model is the easiest level to create, as the modeler is "hiding" all of the business complexity within database concepts such as Entity, Object, and Attribute. If you have no idea how the business works, and you want to cover all situations for all types of industries, an EDM "in the database clouds" can work well.

On the ground. This model uses a minimal amount of business and database cloud entities, with a majority of the entities representing concrete business terms such as Student, Course, and Instructor. This model takes the most time to create of the three varieties. It also can add the most value towards understanding the business and resolving data issues. The model in Figure 1 is at the "on the ground" level.

Table 2 shows each setting value for Abstraction, along with when to choose each setting value. Think about the purpose of your EDM and then select the right setting.

Time

Similar to how the timer on a camera allows you to photograph a future scene, the time setting for an EDM allows you to represent a future to be view on an EDM. An EDM can represent how a business works today or how a business might work some time in the future. The model in Figure 1 can be either a *today* or *tomorrow* view, depending on whether or not it represents the current operations of the business.

Today. An EDM with the *today* setting captures how the business works today. If archaic business rules exist, they will appear on this model even if the business is planning to modify them in the near future. In addition, if an organization is in the process of buying another company, selling a company, or changing lines of business, a *today* view wouldn't show any of this. It would only capture how the business works today.

Tomorrow. An EDM with the *tomorrow* setting can represent any time period in the future. Whether end of the year, five years out, or 10 years out, a *tomorrow* setting represents an ideal view of the organization. When an EDM needs to support an organization's vision or strategic view, a *tomorrow* setting is preferred. I worked on a university EDM that represented an end-of-year view, as that would be when a large application migration would be completed. Note that most organizations who need a *tomorrow* view first have to build a *today* view to create a starting point.

Table 3 shows each setting value for Time, along with when to choose each setting value. Think about the purpose of your EDM, and then select the right setting.

Table 4. Function setting values.

Setting value	Choose this setting value for your EDM when you…
Business	**Need to understand how the business works.** Using business terms, definitions, and rules, how does an organization do business? The business view would show this independent of application terminology and constraints.
Application	**Need to understand how an ERP system works.** Almost all application EDMs are built to understand how a single broad-scoped ERP application works. I started building one of SAP/R3 that was extremely useful in explaining to our department how SAP handles a certain business area.

Function

Similar to how a filter on a camera can change the appearance of a scene, the function setting for an EDM allows you to represent either a business or functional view on the EDM. Are we modeling the application's view of the world or the business' view of the world? Sometimes they can be the same (or very close to the same). Most of the time, they will be quite different. The model in Figure 1 represents a business view.

Business view. This view uses business terminology and rules. The EDM represents an application-independent view. It doesn't matter whether the organization is using SAP or Siebel. The business will be represented in business concepts.

Application view. This view uses application terminology and rules. It's a view of the business through the eyes of an application. If the application uses the term "Object" for the term "Product," it will appear as Object on the EDM and be defined according to how the application defines the term, not how the business defines it. In our survey, there were only two efforts that were at the application level.

Table 4 shows each setting value for Function, along with when to choose each setting value. Think about the purpose of your EDM, and then select the right setting.

Section Two: Use of an Industry Model

EDMs have been around for more than 20 years, yet up until recently you would have no choice but to build your EDM from scratch. Today, however, instead of reinventing the wheel and modeling structures that have been modeled many times before, organizations can leverage industry models. An industry model is a prebuilt data model that captures how an organization in a particular industry works or should work. For example, the Teradata Manufacturing Logical Data Model (MLDM) captures how a manufacturer works or should work.

Use of an industry model is still at its competitive advantage stage and not yet at its competitive necessity stage. A relatively small number of organizations are using them. For example, in our EDM survey, 18 out of 22 (82%) organizations built their own EDM. It's important to understand the factors that can sway an organization to use an industry model, and the factors that can sway an organization to build its own.

Consider the following factors in the context of the purpose of your EDM, and the settings you feel are needed from the prior section.

Factors in Deciding Whether or Not to Use an Industry Model

There are five factors that sway an organization to use an industry model:

Time and Money to Build

If your organization needs to show results quickly, an industry model can give you a head start. Instead of modeling your business from scratch and possibly getting into an analysis-paralysis mode, you can use a model that already exists. The first EDM I worked on was for a telecommunications company. I worked on this model with a team of seven others for two years. The project was around for eight years before I joined. This means we were working on this EDM for more than 10 years, and had not finished! We were trying to build this model from scratch. This is an extreme story, and I haven't seen or read about a similar case since then. However, if you're given six months to complete an EDM, there would be a greater chance of meeting this deadline with an industry model. Using an industry model doesn't magically give you your enterprise view, but it can save lots of time. Instead of spending lots of time modeling, your time will be spent mapping your existing structures to the industry model and discussing with your team customizations to the industry model, such as which definitions should change and which additional data elements should be added.

Time and Money to Maintain

Many EDM efforts fail to consider the routine maintenance required of an EDM. Every application development effort has a two-way arrow with the EDM. It can use structures from the EDM, and it can suggest changes to the EDM. If the EDM is not part of a methodology that keeps it fresh and accurate, it will very quickly become obsolete. An industry model can minimize the risk of this happening by allowing an outside resource to keep the model up to date for you. You still will have to maintain the model with your own organization-specific changes, but when the industry changes, you will have less work to do. For example, when RFID came out, the Teradata MLDM updated its model, and organizations were able to use this without maintaining it themselves.

Risk and Modeling Expertise

Building an EDM requires skilled resources from both business and IT. A business resource must either be very knowledgeable across many departments or very knowledgeable within a single department. It's rare to find business resources with both deep and broad knowledge. I worked at a manufacturing company where there were only three people out of thousands who had this background. Most business resources tend to know one, and only one, area well. This requires numerous business resources to cover the entire enterprise. This leads to the skills required from IT. The more business resources needed to build and then validate the model, the more experienced the analyst must be. The analyst will need to know not only how to extract knowledge from the business resources, but also carefully resolve conflict areas and neatly fit their knowledge together in an integrated view. The modeler, who in many cases is the same person as the analyst, also needs to have advanced modeling skills. It requires modeling for the future and reducing a very complex business into boxes and lines. An industry model lets you leverage someone else's expertise. An analyst at a large semi-conductor manufacturer, which uses the Teradata MLDM, believes there is a huge advantage to getting your model off the shelf: "You are taking the brightest thinkers out there." For example, the Teradata iLDMs use proven modeling methodologies and an experienced professional services team to avoid the common and expensive pitfalls that can arise with building an EDM from scratch.

Existing Application Infrastructure and Vendor Relations

If you currently use a lot of existing tools from the same vendor, and already have a strong relationship with that vendor, you might see if that vendor offers any industry models. The vendor's EDM might work seamlessly with some of the other vendor's products. Teradata iLDMs work with the Teradata Enterprise Data Warehouse Roadmap model product, which creates a very strong connection between the EDM and BI solution, as will be discussed shortly.

Information Sharing within Your Industry

If your business needs to exchange data with other companies within the same industry, an industry model is a good idea. Organizations in the same industry use similar constructs and terminology. For example, within the travel agency, rental car companies and airlines typically exchange information and therefore, using an industry model will allow for the same terminology across organizations. Most industries these days require sharing at least some amount of data across companies.

Which of the previous five factors are important to your organization? Weigh those previous factors that are important with these reasons why organizations build their own, and decide whether or not it makes sense to use an industry data model.

Only need a subject view. If in the detail setting chart mentioned earlier, you've decided your EDM only needs to exist at a subject level, you may not need an industry model. Out of the 18 organizations that built their own EDMs in our survey, eight of them only needed a subject view (44%). Subject models take a relatively short amount of time to create and maintain, and therefore, the time and money savings of an industry model are minimized.

Already have an EDM. If you already have an EDM or a pretty good start to an EDM, it might be easier to use this than use a pre-built one at this stage. Note that if you have several enterprise views or employees with different opinions as to what to name or how to define concepts, you might find that an industry model provides an unbiased external view to which each of the existing EDM views can be mapped.

No suitable industry models exist. Recently, I've been seeing this reason occur frequently. In our challenging economy, organizations are reinventing themselves. This often leads to organizations that cross industries or, at the opposite end of the spectrum, focus on very specific niches within an industry. It would be very challenging and time consuming to customize an industry model to make it work in these cases. An example of an organization that crosses industry lines is an online bookseller that also produces books and advertises products for other vendors. This bookseller therefore is not only a retailer, but also a manufacturer and an advertising agency. An example of an organization in a specific niche is an aluminum-rolling plant where I did some work in the Midwest. There most likely could be some parts of a manufacturing industry model that apply to this plant, but the amount of customization required may not be worth the effort. Note that the iLDM Unification Project, which will be discussed in Section 4, addresses this suitability concern by tailoring an industry data model to a particular organization using a series of data modeling building blocks.

Pioneering culture. If your company believes in building everything itself, the EDM is usually no exception. This pioneering culture is the opposite of the best-of-breed culture, which buys the top-rated products from each category.

Let's look at how a semiconductor manufacturer weighed the factors for and against an industry model. Their reason for having an EDM was to drive an integrated data warehouse. They have a lot of modeling expertise and a pre-existing EDM, which are strong points against an industry model. However, the pre-existing EDM is a collection of physical data modeling structures (not logical) that are difficult to reuse and integrate. The pre-existing EDM, however, is rich in attribution and definitions. Because of the need for an integrated logical view, and because they are currently a Teradata customer, they obtained the Teradata MLDM. They currently use this MLDM as a model template and enhance this template with definitions and additional attributes from the existing pool of physical tables. This gives them the best of both worlds: the MLDM provides the integrated, flexible view, and the existing models provide the richness to the definitions, additional data elements, and business rules. Today there are more than 100 entities in production in their data warehouse that are based on the MLDM.

Industry Model Resources

Industry data models can be found in a number of different places, such as through websites, books, industry organizations, and vendors. Returning to the previous cost/benefit discussion, you'll need to match your cost requirements with the settings from each of these sources.

Websites

You can Google terms, such as "Enterprise Data Model" or "Insurance Data Model," and see if any of the results can be used. Websites work best when you're planning to build your own EDM and want to supplement it or validate it with existing structures. Although this is the least expensive method of obtaining industry models, it could be the most suspect. Many data models on the internet might be outdated, in early release forms, or simply incorrect. However, it can provide some useful information. One time, I needed to understand how different organizations modeled the concept of Person. I found many Person models on the web and was able to combine bits and pieces from many of these to create a robust, new model.

Books

There are several books on the market that could give your EDM effort a head start. Books are an inexpensive and at times a very useful source for an EDM. Books written by Len Silverston and David Hay (see the list of references) could shorten the EDM modeling effort.

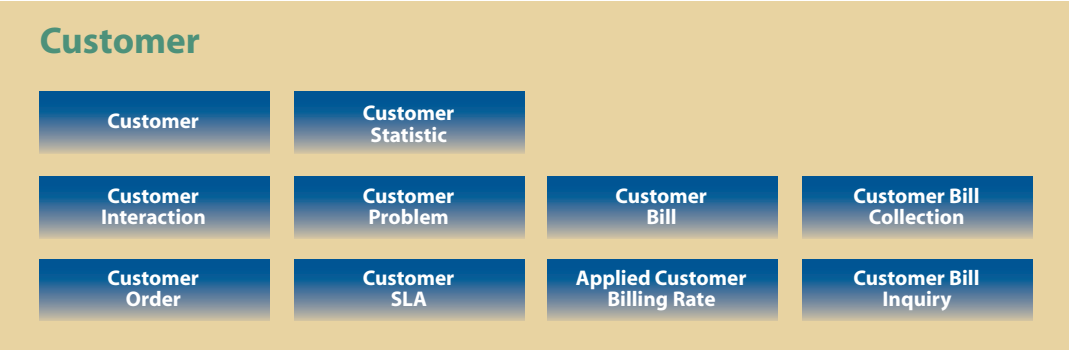

Figure 3. Sample of SID model.

These books provide flexible, reusable structures that you can relatively easily customize for your business. You can also read Chapter 9 from my book, *Data Modeler's Workbench*, if you're looking for some building blocks to build your own EDM. Organizations using books tend to also have experienced data models that are comfortable with abstraction, or they also own industry models to supplement the books.

Industry Organizations

A number of industries, such as telecommunications and healthcare, have a consortium that works together on initiatives useful to all of its members. One of the initiatives that many of these consortiums work to complete is an EDM. For example, the telecommunications consortium has the Shared Information Data Model (SID), and healthcare has the HL7 Reference Information Model (RIM). Figure 3 shows the SID high-level model of Customer and Customer key concepts.

HL7 is an abstract design that offers more detail through subtyping. It's so abstract that you'll find an entity named "Entity," as well as entities named Role, Act, and Participation. However, with the use of subtyping, more concrete terms are incorporated, and there is also a substantial amount of attribution. Links to both SID and HL7 models are included in the References section at the end of this paper. Industry organizations are an excellent source for industry models. However, the number of industries with these consortiums is limited, and they're more useful on the operational side than the analytical side. Their models are often abstract and, usually due to being built by committee over several meetings a year, they tend to evolve slowly over time.

Vendor Models

The most useful, yet also at times the most expensive, industry models are those offered by vendors. There are several companies that offer industry models, with Teradata leading the way. What makes these models most useful is their level of detail and their accuracy. They tend to be "on the ground" models that are heavily attributed. In addition, the modelers who build these models have not only heavy modeling experience, but heavy modeling experience *within a given industry.* The next section will go into more detail about the Teradata iLDMs.

Section Three: Teradata Industry Logical Data Model Overview

Before you use something, you need to understand how it works. You can't use a hammer for example, without understanding the principles of leverage and aim. For BI, this means before you use the data for reporting purposes, you need to understand how the business works. You can't monitor widget sales for example, without an understanding of how to order widgets. *Can a widget appear on more than one order? Can an order contain multiple order lines? Can an order line belong to more than one order?* These are some of the hundreds – and sometimes thousands – of rules that an LDM captures as part of its goal of representing how the business works.

As a global leader in enterprise data warehousing, Teradata has worked with many companies to first understand how their business works and then how to deliver information to make timely business decisions. This leads to a deep and broad knowledge of how a number of industries work. After working with a handful of travel companies for example, you can learn a great deal about the travel industry. Teradata Professional Services has documented this business knowledge in the form of nine distinct iLDMs.

An iLDM is an industry-specific LDM that accurately reflects the way the industry does or will do business. It's a subject-oriented and integrated view of all of the information needed for an enterprise data warehouse (EDW) to answer strategic and tactical business questions. As an LDM, it's application-independent. This means software, hardware, and communications constraints aren't present on the model. All model changes that will be needed for speed, storage, security, backup and recovery, and the like, all need to be added at the physical level. However, a majority of the structures on the logical model do translate nicely into an efficient physical design. It's interesting to note that because the iLDMs are logical (and not physical), organizations can leverage them even if the organizations don't use other Teradata products. However, there are additional advantages to using the Teradata iLDMs, such as using the Teradata EDW Roadmap, that will be discussed shortly.

The structures on the iLDMs are designed for international use, and not just U.S.-based. For example, the term "postal code" is chosen over "zip code." The data elements on each iLDM follow best-practice naming standards, including the use of class words based on the ISO 11179 metadata standard. A class word is the last part of a data element name that represents the high-level category in which the data element belongs. Examples of class words are name, code, identifier, date, quantity, and amount. So the class word for Customer Last Name is name.

Each iLDM is sectioned into subject areas. Subjects are neatly captured in separate views, and the use of color to distinguish each subject area makes it easier to digest the larger models. Even the CDM is color coded by subject area. There are certain subject areas, called Shared Subject Areas (SSA), that are common across the iLDMs. SSAs, such as Party, Event, Geography, and Demographics, have a common core in each iLDM and then are extended where appropriate within each of the models.

The iLDM contains the common information that companies share within an industry and therefore is not meant to be a complete solution for a company. Most companies need to extend the model about 25% through customizations. Customizing includes adding data elements and enriching definitions. Only the key data elements appear on the models, and therefore, any

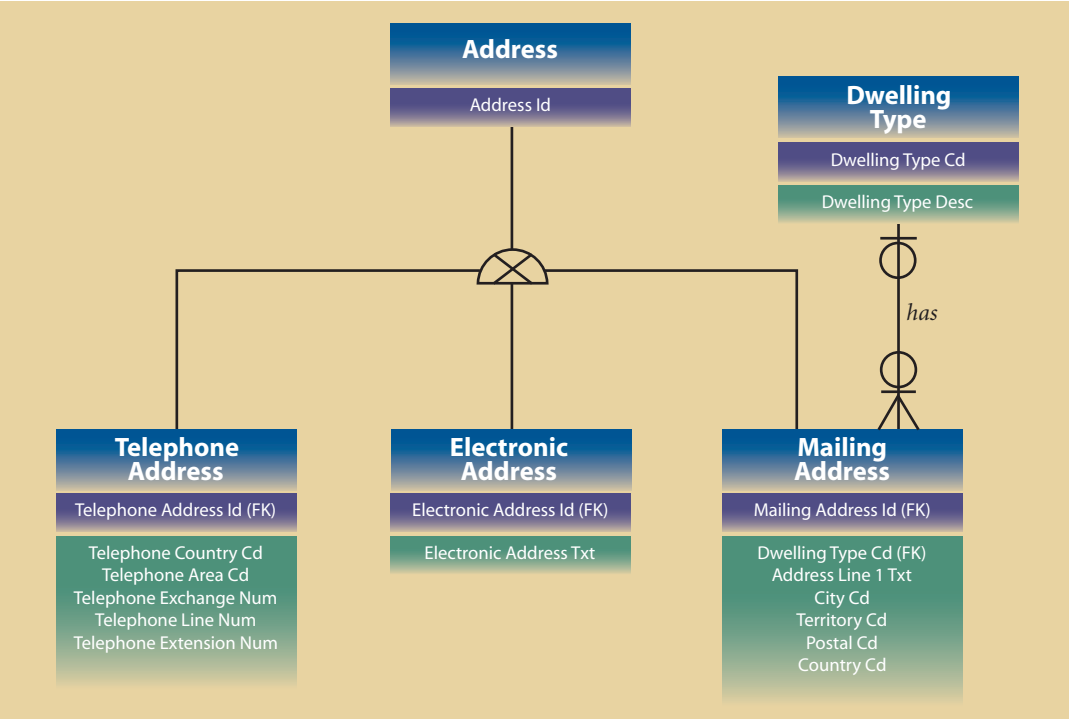

Figure 4. A sample of the Geography shared-subject area.

other data elements will need to be added to the appropriate entity. In addition, in many cases, the starter definitions provided will require additional text to be meaningful to a specific company. Note that it's anticipated that the Unification Project will reduce the customization effort substantially.

The iLDM contains a fair amount of abstraction. Abstraction means combining like things together under generic terms, such as Event and Party. This allows for additional flexibility, as well as additional commonality across iLDMs. All industries have Events, for example, whether they're travel bookings in the travel industry or orders in the manufacturing industry.

To give you a taste of an iLDM, see Figure 4 for six entities that are common across the iLDMs and belong to the Geography shared-subject area. I removed a number of the attributes from these entities and a number of the relationships to other entities from these entities, but the purpose here is to give you a sample of an iLDM.

An Address can be a Telephone Address, Electronic Address, or Mailing Address. There are certain data elements that are unique to each of these subtypes, such as the Telephone Area Cd belongs only to a Telephone Address. A Dwelling Type can belong to zero, one, or more Mailing Addresses, and a Mailing Address can belong to zero or one Dwelling Type. Here is the definition that is provided for the Dwelling Type entity:

Identifies a type dwelling. example: single family, multi-family, etc.

Here is a definition that is provided for the Dwelling Type Cd data element:

A code that uniquely identifies the DWELLING TYPE.

Table 5. EDM settings for iLDMs.

Setting	Setting value	Does the iLDM have this setting?
Detail	**Subject**	Sometimes. Some of the iLDMs, such as the one for manufacturing, come with a high-level model called the Conceptual Data Model (CDM). See Figure 1 for an example.
	Logical	Yes. The iLDM represents a normalized, abstract view of an industry independent of application and technology. See Figure 4 for an example.
	Physical	No. Structures will need to be modified to perform in a given environment. Techniques, such as indexing and views, are popular approaches to modify the iLDM to work efficiently.
Abstraction	**In the business clouds**	No.
	In the database clouds	No.
	On the ground	Yes. There are abstract structures, such as Event and Party, yet along with each of these abstract concepts comes a set of subtyping so that on the whole, the model is more on the ground than in the business clouds.
Time	**Today**	Yes, it can represent how the business works today.
	Tomorrow	Yes, it can also represent how the business might work in the future.
Function	**Business**	Yes, It's a business view of the world.
	Application	No.

Note that both of these definitions will most likely need to be extended to be meaningful in your organization. For example, you might extend the definition of Dwelling Type Cd:

A code that uniquely identifies the DWELLING TYPE, which is a place of residence or a Post Office Box. Places of residence include single family, multi-family, and temporary housing.

Table 5 summarizes the characteristics of the iLDM, using the EDM settings from Section 1.

After deciding that you need an industry model, a good way to determine if the iLDM is right for your organization is to complete a chart such as this for your needs and see if it maps nicely to these iLDM settings. For example, you would not use the iLDM if your goal is to represent an application view of the world.

The actual iLDM is delivered in an ERwin file and three books with corresponding PDF files. ERwin is a data modeling tool that allows user-friendly navigation, viewing, and reporting on the models. The ERwin file contains separate subject areas to view the large

model in manageable chunks. There are also several prebuilt reports for viewing and printing the models and their metadata. Some of the iLDMs, such as the MLDM, also come with a separate CDM ERwin file. The three books include a reference guide specific to the iLDM obtained, the *Unification Model Reference Guide,* and an Appendix listing all of the attributes on the specific iLDM, as well as entity and subject area definitions. (Note that Unification will be discussed in Section 4.)

There are nine Teradata iLDMs:

- Teradata Communications Logical Data Model (CLDM)
- Teradata Financial Services Logical Data Model (FS-LDM)
- Teradata Healthcare Logical Data Model (HCLDM)
- Teradata Manufacturing Logical Data Model (MLDM)
- Teradata Media & Entertainment Logical Data Model (MELDM)
- Teradata Retail Logical Data Model (RLDM)
- Teradata Transportation and Logistics Logical Data Model (Trans LDM)
- Teradata Travel and Hospitality Logical Data Model (THLDM)
- Teradata Utilities Logical Data Model (ULDM)

Communications Logical Data Model (CLDM)

The Teradata CLDM captures how telecommunications, ISPs, cable and satellite companies, and VOIP organizations work. It supports not just traditional telecommunications services, such as SMS and prepaid wireless, but also local number portability, settlements assurance, and e-commerce channels. Data services are also supported, such as video-conferencing, messaging services, and streaming media. It was introduced in 1995 and contains more than 1,800 entities and 7,500 attributes.

Financial Services Logical Data Model (FS-LDM)

The Teradata FS-LDM describes how the retail and commercial banking, insurance, and brokerage industries work. It depicts a holistic view of financial and risk management, covering retail and commercial banking, brokerage, investment, capital markets, charge card, and insurance. Coupled with the EDM Roadmap, it can answer questions about the impact of spending, and key metrics on costing and profitability. It was introduced in 1996 and contains more than 2,700 entities and 11,200 attributes.

Healthcare Logical Data Model (HCLDM)

The Teradata HCLDM was introduced in 2005. It captures how an organization in the healthcare industry works. This is a very broad category that encompasses heathcare providers, health insurance carriers, and pharmaceutical companies. Coupled with the EDM Roadmap, it can help reduce healthcare costs, improve the quality of care, reduce fraud, and ensure compliance with government regulations. It contains more than 1,700 entities and 7,200 attributes.

Manufacturing Logical Data Model (MLDM)

The Teradata MLDM captures how a general manufacturing organization works. In addition, it has structures to accommodate consumer products and durable goods, automotive, energy, and high-tech industries. Coupled with the EDM Roadmap, it can support track and trace, inspections and testing, warranty management, and logistics. Manufacturers can take uniquely identified items and track them through the lifecycle, beginning with production. Track-and-trace analysis can yield benefits around targeted product recalls, warranty expense management, shrinkage control, counterfeit-item detection, and inventory optimization. It was introduced in 2002 and contains more than 1,600 entities and 6,000 attributes.

Media & Entertainment Logical Data Model (MELDM)

The Teradata Media & Entertainment LDM covers the business processes within these two industries, and contains more than 1,100 entities and 4,800 attributes. Some of the benefits of this model include reducing the cost of irrelevant advertising, tracking the best-performing ads, knowing what titles and rights are available for a given territory, and managing subscriptions, inventory, and forecasting.

Retail Logical Data Model (RLDM)

The Teradata RLDM describes how a typical retail store works. It's general enough to apply to a wide variety of retail operations, from department and convenience stores to e-commerce and retail pharmacy. Coupled with the EDM Roadmap, it can strengthen customer relationships, manage merchandise across regions, and track multiple supply chain channels. It was introduced in 2000 and contains more than 1,100 entities and 4,100 attributes.

Transportation and Logistics Logical Data Model (Trans LDM)

The Teradata Transportation and Logistics LDM describes how an organization in the transportation industry works. This includes the transportation, logistics, and distribution industries, such as those organizations involved in trucking, rail, and postal. Coupled with the EDM Roadmap, it can answer questions about the effectiveness of promotional offers, inventory levels, and shipment criticality. It was introduced in 2003 and contains more than 1,400 entities and 5,700 attributes.

Travel and Hospitality Logical Data Model (THLDM)

The Teradata THLDM describes how an organization in the travel and hospitality industry works. This includes airlines, car rental companies, hotels and gaming, air cargo, travel agencies, cruise lines, and passenger rail. Coupled with the EDM Roadmap, it can improve frequent traveler program performance, pinpoint the causes for maintenance and supply issues, and identify the profitability of the business by geography, business unit, and product line. It was introduced in 2001 and contains more than 1,300 entities and 4,900 attributes.

Utilities Logical Data Model (ULDM)

The Teradata ULDM describes how a utility company works. This includes service meter events, billing, deposit events, fraud profiles, customers, suppliers, vendors, human capital management, and payroll. It addresses the challenges of Advanced Metering Infrastructure (AMI) and Smart Grid technology. It contains more than 1,200 entities and 4,900 attributes.

Section Four: iLDM Evolution through Unification

Section 2 discussed the factors that sway an organization toward or away from using an industry data model. One of the common reasons given for organizations swaying away from using an industry data model is that "no suitable industry models exist." In today's economy, where there is a tight relationship between creativity and survival, many organizations perform functions that cross industries. The retailer that also manufactures goods, the airline that brokers rental cars, and the phone company that sells cable and internet services are all common examples of cross-industry organizations.

As organizations continue to diversify by expanding into different markets and different industries, a particular iLDM might require substantial customization to make it fit an organization's enterprise. Some organizations may need two or more iLDMs to represent the enterprise and therefore either integration of the iLDMs is needed before the model becomes useful, or heavy customization to the iLDM is required.

For industry models to remain useful, they need to accommodate our changing organizations. One impractical approach is to offer industry models that represent every flavor of every organization. However, the amount of effort to maintain all of the combinations would quickly become overwhelming.

There is another approach, however. Instead of building up industry data models with all possible combinations, integrate them through their common components. Leverage the similarities across industry data models and create building blocks that allow organizations to pick one primary iLDM and then add on the relevant common component building blocks to create their unique models. This is the solution Teradata has created with its unification project.

Unification is the integration of all iLDMs creating a plug-and-play industry data model environment. In addition to having separate iLDMs for each industry, each iLDM offers building blocks to the other iLDMs. Roughly 45% of current iLDM models are candidates for unification, and therefore make up the building blocks. About 55% remain specific to an industry. An organization still selects one of the iLDMs to represent its primary business. Along with this iLDM comes a reference guide containing all of the building blocks called "independent modules."

Advantages of Unification

There are a number of benefits from unification including:

More frequent model releases. Instead of receiving model releases annually, releases can be sent whenever a building block is enhanced.

Less customization. By adding the necessary building blocks to an existing iLDM, there is less model customization needed to fit a particular implementation.

Greater applicability. As organizations continue to morph across industries, the iLDMs can accommodate this.

Foundational and Functional Unification Building Blocks

Unification building blocks fit into two categories: foundational models and functional areas. Foundational models represent the who, what, when, where, why, and how that cross all industries. For example, the fact that a Party can play many Roles is true regardless of industry. These foundational models are the glue across the models. The functional areas are 27 common business processes, such as call center processing and category management.

Table 6 captures the foundational model names and the key functional areas names. The intersection cells represent examples of what a particular iLDM customer would call a foundational concept within the functional business process.

Table 6. Foundational model names and key functional areas names.

Functional	Foundational					
	WHO (Party)	**WHAT (Product/ Service)**	**WHEN (Time)**	**WHERE (Geography)**	**WHY (Event)**	**HOW (Document/ Agreement)**
Call Center Processing	Employee Consumer	Item Component Feature	Time of Initiation Time of Duration	City Placing Call Location Handling Call	Consumer Interaction (phone call)	Trouble Ticket Order Complaint
Web Interaction	Customer	Item	Time of Initiation Time of Page Entry	IP Address (actually used in fraud analysis)	Web Visit	Sales Order Delivery Order
Shipment	Shipping Party Carrier Party	Item	Time Ordered Time Picked Up Time Delivered Time Delayed	Pickup Location Delivery Location Stopoff Location	Shipping Order	Bill of Lading Measurement Events (temperature, shocks)
Promotion	Authorizing Party Media Party	Items Multi-Media Components	Time Ordered Time Placed	SMSA Area of Dominant Market Influence	Ad Placement Campaign Plan	Ad Order
Profitability Analysis	Customer Delivery Organization	Item Sale Price	Time of Sale Time of Production	Storage Location Delivery Location	Customer Interaction	Sales Order Production Order

Section Five: Areas that Benefit from the EDM and iLDM

An EDM and industry models, such as the iLDM, provide organizations with a single integrated view of their business information. Integration is very difficult for many organizations to achieve, yet it's a prerequisite for most organization-wide initiatives. In fact, according to a recent TDWI Flashpoint, more than 69% of organizations surveyed find that integration is a high barrier to new application development. However, if an organization can integrate its data, the rewards are great. Continental Airlines, which in 1994 was ranked lowest by U.S. Department of Transportation metrics, has recently been named the "most admired global airline" by *Fortune Magazine.* This turnaround is due to Continental obtaining an integrated view of more than 25 of their source systems. This enterprise view has not only boosted customer service, it also has saved Continental more than $25 million and brought in more than $150 million in additional revenue.

To give you a small taste of EDM integration, assume part of your EDM is the iLDM Geography structure from Figure 4. Along with this model comes a spreadsheet mapping each of the existing places the field is stored to the EDM structure. See Table 7.

There are three data elements in this example that map to the EDM Electronic Address Txt. Even with such a simple example, you can see some of the complexities, such as in resolving the same address that was entered in multiple locations or in parsing out the valid email addresses from the company data element.

The EDM, along with a mapping back to existing systems, offers the greatest benefits for organization-wide initiatives, such as ERP, master data management, BI, data quality efforts, and training.

Table 7. Email address mapping.

Source field	Source table	Source system	Create, read, update, or delete?	Comments	EDM field	EDM table
email	Associate	HR Payroll	Create	Straight map	Electronic Address Txt	ELECTRONIC ADDRESS
e_mail_addr	Vendor	Supplier Contact	Create	Straight map		
company	Consumer	Consumer Contact	Create	90% of time this is the email address and 10% it contains the company name. Search for @ to ensure you have the email address.		

Enterprise Resource Planning

ERP is an information system that integrates all functional areas for an entire enterprise. Broad areas, such as manufacturing, human resources, and finance, can all be managed by ERP systems, such as those offered by SAP and Oracle. Because ERP packages cross broad areas, there is a need to view information at an integrated subject level, and therefore an EDM becomes an invaluable tool. Two organizations in our survey (9%) built an EDM to assist with an ERP initiative. An EDM can assist ERP efforts in two main areas: evaluating whether a particular ERP package is a good fit for an organization and performing both internal and external impact analysis to the ERP package.

An EDM is a great tool for determining whether a particular ERP package would be a good fit for your organization. ERP vendors should provide some type of model of their system along with definitions, and an analyst can map the ERP system to your EDM to identify any potential data issues prior to investing time and money in the package. A subject-level EDM matched against a subject-level ERP model, for example, can highlight to management on a single page what the pros and cons are of different ERP packages in terms of organization fit. Coloring entities with green (representing a good fit) and red (representing an issue area) can be a simple, effective communication tool to gauge whether the ERP package should be pursued. If your EDM is also mapped to your existing applications, it can provide insight into what systems will be replaced by the ERP package, and which systems will need to receive data from or send data to the ERP package.

Impact analysis is a frequent use of the EDM for ERP. I participated in an SAP rollout for a global manufacturing company. There were several hundred business and IT employees working hard to ensure the rollout was a success. However, despite the enormous amount of testing that was done within each of the SAP rollout departments, several unanticipated issues occurred. An SAP report, for example, was updated by two different departments without knowledge that each department was updating it. I built a high-level model of part of SAP/R3 for this company to help identify touch points and overlaps within SAP rollouts. It also became a very good communication tool for describing how SAP interacts with other systems and functional areas.

Teradata offers customers a mapping between the MLDM and SAP/R3. This means if your data warehouse will receive data from SAP, the mapping between SAP/R3 Idoc and the EDM can be invaluable to reduce the amount of time and increase the accuracy of loading SAP/R3 data. A large semiconductor company has taken advantage of this mapping and is reaping the benefits.

Master Data Management

Master data is all of the data that is shared outside of a single application. Most of the time, master data refers to reference data such as Customer and Product, but it can also include event data such as Orders and Credits. Any data that is of use to two or more applications is considered master data. Fourteen organizations in our survey (64%) built an EDM for master data.

An EDM forces a single set of metadata for each data element. This metadata includes a single name, definition, format, domain, and set of business rules. Defining all of this data at an organization level makes master data management possible. The work of creating the organization-level information is already done. In addition, because most EDMs also map to actual applications, master data management identifies which applications use which EDM structures. It becomes an integration point among applications. For new applications, the EDM provides a starting point for their modeling efforts.

"The EDM is a key tool that will help us make sure that the warehouse delivers its promised return on investment. It's the component of data architecture that supports the mission of the data warehouse, to provide business intelligence and enable analysts to form tomorrow's profit-making strategies."

Kathy Long, Knowledge Partners, Inc.

"For enterprise data initiatives, such as an Operational Data Store (ODS) or Data Warehouse (DW), an EDM is mandatory, since data integration is the fundamental principle underlying any such effort. An EDM facilitates the integration of data, diminishing the data silos, inherent in legacy systems."

Noreen Kendle, Delta Air Lines

Business Intelligence

Consider the two quotes at left from analysts who have built EDMs for BI.

The Data Warehousing Institute defines BI as "The processes, technologies, and tools needed to turn data into information, information into knowledge, and knowledge into plans that drive profitable business actions. BI encompasses data warehousing, business analytic tools and content knowledge management." The data warehouse is responsible for this first part of the definition: turning data into information. To turn data into information requires both a subject-oriented and integrated view. These are also the two main ingredients of the EDM and therefore, the reasons the data warehouse needs the EDM. In fact, the integration characteristic is so important to BI that Gartner Group said way back in 1990 that the lack of data integration across information systems makes it difficult or impossible to answer cross-functional or cross-divisional questions. Eighteen organizations in our survey (82%) built an EDM for BI. BI is the most popular reason for building the EDM.

I worked with a university in California to build a subject EDM. The purpose of the EDM was to gain campus-wide agreement on terms, definitions, and business rules so that the organization can intelligently decide which area to focus on first for their data warehouse.

The iLDMs provide the underlying structure necessary for the successful implementation of an EDW. This is because the iLDMs are used in conjunction with another Teradata product, called the EDW Roadmap, which contains a collection of pre-built information needs that are specific to an industry. Needs in the form of improvement opportunities, business questions, and key performance indicators (KPIs) are mapped to the corresponding EDM structure. For example, the Communications industry EDW Roadmap has as its underlying source the Communications iLDM.

By mapping an iLDM to your existing set of applications, and by using the EDW Roadmap, you can easily answer many business questions. This combination supports everything from straightforward, but often difficult to answer questions, such as "who are my most valuable customers?" to queries and analysis demanding data integration among departmental silos, such as analyzing the impact of operational events on revenue and customer behavior. For example, in the Travel iLDM, here are some sample questions that can be answered by combining the two:

- Who are my high- and low-value customers?
- Who should I target to maximize response and avoid revenue dilution?
- From which products and markets should I exit? What is the impact? To which customers?
- Which of my sales and distribution channels are the most effective for each of my target customer segments?

Many organizations obtain the iLDM simply because of its connection with the EDW Roadmap. The questions that can be answered "out of the box" with the EDW Roadmap are important industry-level questions. It can really bring your BI solution up to the next level, and add substantial business value. In fact, coupling the EDM with the Roadmap is the quickest way to EDM ROI.

Data Quality

An EDM exposes data quality issues through its mapping to existing applications. It can show, for example, customer information can be updated in more than one application or product information being retrieved from an incorrect source. Recall, for example, the issue from Table 7 where the email address data element is updated by multiple sources. Besides showing existing data quality issues, it can also prevent future data quality issues by encouraging all new developments to leverage the EDM. Thirteen organizations in our survey (59%) built an EDM for data quality.

Also, it can prevent data quality issues through serving as a communication medium across departments. Folks from different departments can use the EDM as a communication medium as all of the subjects and their definitions and relationships are consistent across departments. If issues arise, they can use the EDM as a map to bridge misunderstandings and complex data anomalies.

In addition, it can help bring issues to the surface during the process of building an EDM. In one particular example, while building an EDM, participants with lots of manufacturing business experience started debating the differences between an ingredient and raw material that could have led to substantial application issues later on in the process.

Employee Development

An EDM is a great training tool. This is because if someone is new to the industry, company, or department, they can come to one place to get an understanding of how the company works. Our industry requires employees and consultants to come up to speed very quickly and start contributing to a project, and an EDM is a good way to quickly grasp how an organization works.

One financial company in our survey (5%) built an EDM for training. The analyst says this about the EDM: “I have used it as part of a new-hire orientation. I call it 'Where's the Data?'.”

Section Six: Use of the iLDM

I was asked to spend several days in Europe working with a large telecommunications company to come up with an approach and start building its EDM. I jumped at this consulting opportunity, as the phrases "Europe" and "Enterprise Data Model" both sounded extremely exciting. They had access to an industry model, but weren't sure of the best approach to using this model. We divided the group up into two teams and gave each team complete freedom to work the industry model into their methodology and approach. After about an hour we came back together as a group to review our initial work. We were all amazed how completely different the two approaches were. One group recommended building its own EDM from scratch and using the industry model as a model to ensure there were no gaps and that the correct terminology was used. The other group used the industry model in its entirety and from the beginning, and then looked at existing structures to ensure each data element mapped to the enterprise model. This section will discuss the approach in building the EDM and how to leverage the iLDMs in the best possible way. We will also discuss how you can obtain a quicker ROI with an iLDM.

Approach

Regardless of whether an industry EDM is used, these four steps need to be addressed:

1. Identify EDM objectives and stakeholders.

2. Build the business solution.

3. Build the technical solution.

4. Maintain the EDM.

We will discuss each of these steps in the following sections. Building the business solution is where the iLDMs fit. There are three ways organizations can use the iLDM in building their business solutions.

Identify EDM Objectives and Settings

Refer to Figure 2 and the individual EDM settings, and match your organization's reasons for having the EDM with the appropriate settings. For example, if you're short on time and need to get a relatively quick consensus about how the current business works today, you most likely will choose an "on the ground" subject model showing the current view of the business. If you have more time and need to create a detailed integrated view for your EDW, you most likely will choose an "on the ground" LDM showing the current view of the business. As Stephen Covey would say, it's extremely important to "begin with the end in mind." Make sure there is agreement about why you're building an EDM and the appropriate settings that would meet your EDM's goals. There is no one setting that is always right or always wrong. It depends on what you need.

I experienced a taste of what not to do on my first EDM effort. We were aiming towards a telecommunications EDM, and after many years of detailed logical data modeling, we lost our funding. Looking back on this experience, one of the main reasons this project failed was that we never had an agreed-upon objective between business and IT about why we are building the EDM in the first place.

Produce the Business Solution (the Subject and Logical Data Model)

The business solution is a view of the business independent of any physical constraints. It captures all of the concepts and rules that are important to the enterprise. An industry model is useful at the business solution stage. Each business has its own unique software and hardware infrastructure, and therefore, also its own unique physical EDM. However, the subject or logical EDM that captures the way the business works tends to be more common across companies within the same industry.

Out of our EDM survey, only four out of 22 companies (18%) used industry models to help with the EDM effort. Of these four organizations, currently only two use the iLDMs. I believe that as an awareness of industry models and iLDMs increases, and as the need for the EDM increases, the number of companies that use iLDMs in building their EDMs will continue to grow. iLDMs tend to get used in one of three ways during the EDM process:

- As the EDM
- As an input to the EDM
- As a tool to identify gaps

iLDM is the EDM

Using the iLDM as the EDM means that, with the exception of some model customization, the organization will completely use the iLDM. Any existing models will need to get mapped into the iLDM. Even though existing systems might look very different from the Address structure in Figure 4, this Address structure will be used, and existing systems will be mapped directly to it. The iLDM becomes the system of record for the EDM. With the unification project, this option becomes more of a possibility.

iLDM is One Input to the EDM

Using the iLDM as an input means that there are a number of enterprise-type models already existing in the organization, and the iLDM is another model that needs to be integrated. There are many different degrees to which the iLDM can be relied on. It can be the focal point for all of the existing models to map to, or it can be just another model to be used. The Address structure in Figure 4 becomes one of the inputs to the EDM, along with one or more other models. Each of these models has to be reconciled to create a new model that combines the best of each of the models.

Another example, a large semiconductor manufacturer uses the Teradata MLDM as the primary, yet not the sole input, into their data warehouse data model. They use the MLDM as the integrated source, and then also consult the repository of models to identify additional data elements they might need. They also consult the repository to find richer definitions, as the definitions in the MLDM are written to be generic starter definitions.

When an iLDM is an input to the EDM, you might decide to rely more heavily on the CDM. The CDM provides a high-level overview to the iLDM and is about 10% of its size. For example, the MLDM has slightly more than 700 entities, and the manufacturing CDM has about 70 major entities. It only shows the most important entities in an iLDM and their relationships. It's a good way to get a high-level overview to a given organization, and if you find yourself sewing existing EDMs together into one current and complete picture, the CDM could provide a high-level roadmap to this patchwork quilt.

iLDM is Used to Identify Gaps

This is the least reliance on an iLDM, yet still an important role that the iLDM plays. Using the Geography example, an organization already has a fairly robust address structure that is part of their EDM. However, they use the iLDM as a way to identify missing data elements and relationships on their model. They also use the iLDM as a way to see information in a different way. For example, perhaps they did not view a telephone number as a type of address, but maybe it makes the most sense to do so.

Produce the Technical Solution (the Physical Data Model)

The technical solution is the business solution modified to perform in a given environment. Software, hardware, and communications are all factors that need to be considered on the physical data model. The model is adjusted for these factors to enable quick performance, efficient use of storage, and several other factors, including security and error recovery. Techniques that are used include the use of database views, denormalization, indexing, and partitioning.

The iLDM will need to be modified because it represents a business solution and not the technical solution. Teradata recommends the use of database views over partitioning if modifications are needed. This is because the views sit over the normalized business structures enabling greater flexibility and referential integrity.

Maintain EDM

The maintenance step is the most overlooked step of the four and perhaps the most important. As soon as an EDM turns from a proactive role to a reactive role, often the model doesn't get kept up to date and loses its value quickly. In our survey, a large number of respondents mentioned that their EDMs aren't kept up to date, and this had a direct impact on value. Here are a few verbatim responses:

- "It's not kept up to date and doesn't play a role in software development."
- "It was at a very detailed level and was near impossible to keep up to date."
- "Unfortunately, my duties were expanded outside of data management, and the EDM is no longer maintained."

Those companies that use the iLDMs have a definite advantage when it comes to maintenance. This is because each of the iLDMs is maintained, and updates are available at least annually. For example, organizations using the MLDM did not have to model RFID when it became available, as the next release of MDLM already had this included.

Quicker ROI with an iLDM

Jan Kamil, senior data modeling consultant, and Yvonne Balditt, data analyst, proposed a detailed calculation of EDM in the May 2006 DM Review Design Challenge. "Assume EDM ROI = (EDM contribution to profit) – (cost of the EDM). The contribution to profit can include factors, such as the cost savings of staff time not spent on research because an EDM exists, and a reduction of outside marketing costs because the data to support customer trends, such as retention and attrition, will eventually exist in the EDM's holistic customer view. For example, you can save $15,000/year paid to an outside marketing company for information that now exists within the EDM. The cost of the EDM asset includes not only staff wages, hardware, and software, but also lost opportunity costs (the cost of the ROI of projects that staff is not working on because they're building the EDM)."

It's very interesting that despite the large amount of time and financial effort required by an EDM, few organizations quantify its benefits in terms of ROI. I am guilty of this as well, as there have been several EDM efforts I've worked on where the benefit was not apparent – or at least not quantified in terms where management and the business could easily see value. For example, here are a few quotes from our EDM survey on ROI:

- "I couldn't tell you."
- "I don't know."
- "Blank check."
- "Not sure."
- "I didn't keep track of time."

Some organizations, on the other hand, keep a very close watch on the time and money, illustrated by these quotes from our EDM survey:

- "Approximately eight-person months for the first version."
- "1 year, $ .5 million consulting, plus internal costs for business and IT support"
- "It has been in the works for 2+ years (calendar time), about 200 person hours total so far."
- "Two years and more than $1 million."

In terms of value, the actual ROI can be more than just monetary savings, as shown by these comments from our EDM survey:

- "It was on everybody's cubicle walls."
- "Even people who were very critical in the department agreed with the model and definitions by the end of the session."
- "The EDM is in its infancy, however, the EDM has been used to scope two enterprise efforts and endorsed by senior IT management."
- "We don't have a quantitative measure. Success is measured by how much it's used across projects."
- "Our EDM is widely used and accepted by business units, and they came back for more data from us."
- "ROI is not used to make IT-related investment decisions at this company, at least not in any formal way. Its success can be measured on how much the model is used by various application teams, and whether it properly represents the business process/data integration."
- "I took the presentation on a 'road show,' and people were excited by it – saw where we had redundant processes. Upper management took a real liking to it because it revealed where we could save money by reducing redundant data processing."

A senior marketing manager for Teradata Professional Services, said: "There's no reason to spend time and money starting from scratch when best practices exist that can be tailored per individual organization. By mapping prioritized business goals to a consistent data model, we're able to shorten the ROI cycle." It takes less time on average to use something existing rather than recreate it, and this, coupled with how well the iLDM leverages other Teradata tools, such as the EDW Roadmap, allows for quicker EDM ROI.

Section Seven: EDM Advice

An EDM is needed by almost every corporate-wide initiative, and an industry model can facilitate the task of creating an EDM for your organization. Use Section Five in this paper to identify one or more initiatives as the driver behind your EDM. Then use Section One to determine, based on these initiatives, which EDM settings are most appropriate. Then use Section Two as a guide to determine if an industry model is right for your effort, and follow it up with Section Three to see if the iLDMs make sense. Leverage unification in Section Four, and reduce the amount of customization an industry model typically requires. Use Section Six to create a project plan for building the EDM using the four major steps as a guide, including planning for which of the three iLDM usages are right for your organization.

Responses to our EDM survey produced useful advice from people who have built and are using EDMs in their organizations. Here is some of this advice in the form of a "Top 10" list.

1. The first step is to identify the purpose of the EDM. "Have clearly defined goals/objectives and expected outcomes for the EDM in terms of its use in the rest of IT."
2. Leverage existing modeling work. "Buy as much as possible, then use this to jumpstart the effort. Business analysts in a company do much better with a straw-man than a blank sheet of paper. Industry-standard LDMs also deliver knowledge from other companies in a neutral way, which may not have been considered in the particular situation."
3. Maintain, maintain, maintain. "Need a process in place to ensure that it's still enhanced and maintained even during crunch times. That is, as soon as there were shorter time priorities, the EDM got lost in the shuffle."
4. Take small steps. "Start small, with the basic subjects. Include definitions. Share it as much as you can, and be open to ideas and suggestions so that more people have buy-in." Another respondent replied with this visual: "Cut the elephant into bite-sized pieces, and create small successes rather than a monolithic failure. That will also create greater buy-in."
5. Get a business sponsor, and make sure you always have one.
6. Have talented, enthusiastic resources. "Get a GOOD data architect, technical communicator, and business analysts." Someone else also said, "Get an external consultant who has done it successfully."
7. Educate, educate, educate. An understanding of data management needs to be constantly nurtured. It also needs continuous management support.
8. Worth the effort. "It's a long and painful process, but worth every bit of time and money spent – just like childbirth."
9. "Don't assume it has to be complete to be useful."
10. The process of creating an EDM is just as important as the end result. "A large part of the value of the model was obtained in large meeting rooms with more than 40 people present. It really helped everyone get on the same page."

About the Author

Steve Hoberman is one of the world's most well-known data modeling gurus. He understands the human side of data modeling and has evangelized "next-generation" techniques. Steve taught his first data modeling class in 1992 and has educated more than 10,000 people about data modeling and business intelligence techniques since then. He is known for his entertaining, interactive teaching and lecture style (watch out for flying candy!), and is a popular, frequent presenter at national and international industry conferences. Steve is a columnist and frequent contributor to industry publications, as well as the author of *Data Modeling Made Simple, Data Modeler's Workbench,* and *Data Modeling for the Business*. He is the founder of the Design Challenges group and inventor of the Data Model Scorecard®. You can email Steve at **me@stevehoberman.com**.

References

"Design Challenge: What is the Enterprise Data Model ROI," Steve Hoberman, DM Review, May 2006.

"The Enterprise Data Model: A Key Ingredient for Successful Data Warehousing," Kathy Long, *The Data Administration Newsletter* (www.tdan.com).

"The Enterprise Data Model," Noreen Kendle, *The Data Administration Newsletter* (www.tdan.com).

Hay, D., *Data Modeling Patterns.* Dorset House Publishing Company, Incorporated, 1995.

Hoberman, S., *The Data Modeler's Workbench*, New York: John Wiley & Sons, Inc., 2001.

Hoberman, S., *Data Modeling Made Simple, 2nd Edition.* New Jersey: Technics Publications, LLC, 2009.

Hoberman, S., Burbank, D., Bradley, C., *Data Modeling for the Business.* New Jersey: Technics Publications, LLC, 2009.

Marco D., Jennings M., *Universal Metadata Models.* New York: John Wiley & Sons, Inc, 2004.

Silverston, L., *The Data Model Resource Book*, Revised Edition, Volume 1, A Library of Universal Data Models For All Enterprises. New York: John Wiley & Sons, Inc., 2001.

Silverston, L., *The Data Model Resource Book,* Revised Edition, Volume 2, A Library of Universal Data Models For Industry Types. New York: John Wiley & Sons, Inc., 2001.

Simsion G., Witt G., *Data Modeling Essentials, Third Edition*. San Francisco: Morgan Kaufmann Publishers, 2005.

More information about the HL7 Reference Information Model (RIM) can be found at hl7.org.

More information about the SID (Shared Information Data Model) can be found at www.tmforum.org.

Raising Intelligence is a trademark, and Teradata and the Teradata logo are registered trademarks of Teradata Corporation and/or its affiliates in the U.S. and worldwide.