January 2010

Fast-Tracking Data Warehousing & Business Intelligence
Projects via Intelligent Data Modeling

Claudia Imhoff, Ph.D

Sponsored by:

embarcadero®

Table of Contents

Introduction

Why do we read about so many Business Intelligence (BI) projects missing the mark with their business users? Did the implementers misinterpret the requirements? Were the users unable to verbalize their needs accurately or clearly? Was the wrong technology deployed?

Probably none of these caused the mismatch. Often it is simply a matter of not properly documenting the needed data and related business rules. In other words, the root cause is a poorly constructed or the complete lack of a data model for the application and therefore, no way to confirm the requirements as stated by the business users.

At the core of any BI should be the ability to align business needs with the data infrastructure supporting them. This is almost impossible to do without a data model. Yet many BI implementers do not understand the need for or the benefits from these design components. This paper will examine the major benefits that data models have on BI environments.

What is a Data Model?

First let's define what a data model is. A data model is a "blueprint" of the data infrastructure for an application or environment. This blueprint translates business concepts into technical diagrams that record the fundamental aspects of the information needed for a particular BI environment. In IT application development, there are different levels of data models, depending on the audience for these models. Figure 1 depicts these different levels of models. These data models are used as specialized communication devices so that the business and IT can fully comprehend the BI environment's requirements for data and the business constraints on that data.

There are four levels of data models – each successive level inherits the information and constraints from the previous level but adds specific information required at that level. The first two levels of data models are enterprise standards and should be used as a starting point for any application. Therefore, they are labeled as "Program Activities". The last two levels are specific to a coordinated set of projects under the program. An example would be the set of projects that create the BI environment. These models are labeled as project-specific activities.

- Subject Area Model – A collection of the high-level "subject areas"

and their relationships. The subject areas are groupings of things of importance to the enterprise and usually number between 12 - 15 subject areas.

- Enterprise Data Model – A representation of information used in an enterprise from a business perspective. This model is also called the Logical or Business Data Model. It is designed to be independent of functional application or physical implementation. There is only one enterprise data model in an organization.

- System Model – A collection of the information being addressed by a specific system or application. It is the electronic representation of information and is independent of the specific technology to be used in its implementation. It is developed from a subset of the enterprise data model. Since a system or application may be implemented on multiple technologies (e.g., a data warehouse plus several dependent data marts), there may be multiple system models.

- Technology Model – A collection of the information being addressed by a system and implemented on a specific platform. The technology model must take into account and document the specific hardware, operating system, DBMS, etc., requirements. Just like the system data model, there may be multiple technology models. That is the same system may reside on multiple technologies (e.g., a data warehouse that is physically implemented on different technologies)

Figure 1: Data Model Levels
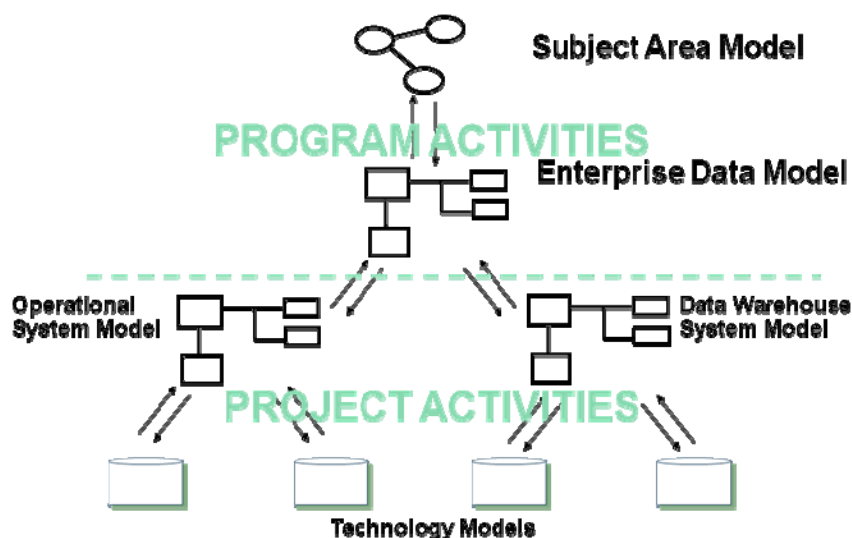
Benefits of Data Models on BI Projects

So what are the benefits of using these data models?

1. *Documentation* – As stated, data models are wonderful communication devices. Just as architectural blueprints are created with several levels of detail, so are data models for the BI environment. Each level produces a successive level of detail about the data. Architectural blueprints record the high level layout of the overall building all the way down to the detailed plumbing and electrical wiring for each room. In the same manner, the data models for the BI environment *document* the overall high level data requirements all the way down to the technical formats and constraints on each individual data attribute. This documentation is invaluable for ensuring that the BI application is correctly supported by the data infrastructure.

2. *Business rule adherence* – Architectural blueprints also include constraints for the building (types of pipes and wires needed, how far apart components must be for safety, structural bearing walls, etc.) Data models also provide a similar record of all the business rules the enterprise must not violate. These can be as simple as an order must have a relationship with both a customer and a product to very complex rules surrounding compliance requirements or proper accounting procedures. These rules of the business are generally captured in the relationships between business entities such as between the customer, product, and order or within the domains of entities, their subtypes, and ultimate formats. Business rules are the checks or controls within the application that stop ETL processes from incorrectly loading the data to restricting analytic processes (or analysts!) from creating invalid comparisons, analyses, or aggregations in the BI environment.

3. *Productivity* – A well-known best practice for BI environments is for each successive project to build upon the foundation created by the previous projects. In other words, the ETL processes, integrated data, and especially the data models should be reused where possible. Well-constructed data models are significant productivity tools that eliminate redundant design efforts, inconsistent data elements, inaccurate or incompatible business rules, etc. They promote significant increases to overall project productivity. We all

know that we tend to be better editors of an existing component than creators of something new. What better way to increase productivity than to start with an existing data model rather than a blank sheet of paper!

4. *Easier maintenance/enhancement* – No matter how thoroughly you investigate the requirements for a BI application, I can guarantee that once it is deployed, the users of the application will figure out new and different ways of using it. This generally translates into enhancements or maintenance to the underlying data infrastructure and processes. Without a data model to act as a roadmap for these changes, they can be very detrimental to the overall performance and usability of the application. For example, the data model determines the optimal placement of new attributes or entities for the best possible response time and ease of use.

5. *Enterprise-focus* – As mentioned in number 3, each BI project should build upon existing deployments. Another best practice is to ensure that the data models for each project maintain the enterprise view or at least highlight any exceptions to this view. The enterprise nature of the data models ensures that the data infrastructure can be reused by multiple departments or groups as well as multiple projects with minimal effort. It may mean that the company has to resolve its differences for such hot entities as customer or product – just agreeing to their definitions can be challenging – but the consistency and reliability obtained by this effort far outweighs the effort it takes to consolidation these differences. In the long run, the overall BI environment benefits in terms of data consistency and reliability.

Getting Started

If a subject area model or enterprise data model do not exist, then a best practice for BI states that the first BI project should create the subject area model (it is a very quick model to create) and only that portion of the enterprise model that directly supports the current BI project and no more. Each successive BI project can then build out more and more of the enterprise data model as the project creates its own system and technology models (see the back arrows in Figure 1). This practice of iteratively building out the two program models means that no one project is responsible for the creation of the entire enterprise data model (a significant effort to say the least). The risk is that the program models may undergo significant changes as successive

projects are implemented. These changes could impact the already implemented projects' models as well.

Another best practice for getting started is to start with the database schema of the existing operational or transaction (source) systems. It is possible to convert these designs into technology and system models. These can in turn be used as a starting point for the enterprise data model and subject area model.

They may also serve as the starting point for the system and technology models of the data warehouse and/or data marts. Reusing their definitions and documentation can be a great leg up on the creation of all the levels of models and can ensure consistency from one environment to the other. But ensure that you have other sources of information for the data models in addition to the operational models, especially if your operational systems are packages like ERP systems. It may be that the business actually wants to change some of the rules embedded in the operational models but can't. Don't continue this frustration by continuing the poor business rule in the BI environment when it is not necessary.

In any case, watch out for analysis paralysis. A data model is never really finished; it can always be polished and polished to the detriment of any project's time line. To avoid this over-design situation, you may want to time box the data model development activities. When the time is up, go with the model that you have. It may not be "perfect" but it will be good enough.

Summary

Given these major benefits of data models, it should be obvious that their usage translates not only into far superior BI applications and environments but also into significant cost reductions by eliminating data inconsistencies and design duplications. Data models also reduce costs by improving the maintenance of BI products and overall efficiency of deployments.

Data models promote a common understanding between IT and the business and are the keys to deploying more successful projects. The likelihood that the BI environment will be adopted and used by the business community goes up significantly when both the business and IT have a thorough understanding of its infrastructure. And that is the whole point of a BI environment – happy and satisfied business users making accurate and timely decisions for the enterprise based on reliable, documented information.

About Claudia Imhoff

President, Intelligent Solutions Inc.

A thought leader, visionary, and practitioner in the rapidly growing fields of business intelligence and customer focused-strategy – Claudia Imhoff, Ph.D., is a popular and dynamic speaker and internationally recognized expert on analytical CRM, business intelligence, and the infrastructure to support these initiatives – the Corporate Information Factory (CIF). Dr. Imhoff has co-authored five highly-regarded and popular books on these subjects and writes monthly columns and articles (totaling more than 100) for technical and business magazines. She has served on the Board of Advisors for DAMA International and was chosen by the DAMA organizations to receive the 1999 and 2005 Individual Achievement Awards. She is an advisor and a faculty member for The Data Warehousing Institute and serves as an advisor for several technology and commercial companies. Dr. Imhoff delivers keynote addresses at conferences sponsored by software companies and their user groups, The Data Warehousing Institute, The Economist, COMDEX, and many international organizations. She has appeared repeatedly on World Business Review, Microsoft's Getting Results programs, and web casts sponsored by DM Review, Better Management, and several technology vendors. She is a member of the Advisory Board of the Daniels School of Business at the University of Denver and is on several technology companies' advisory councils.

Dr. Imhoff founded Intelligent Solutions, Inc. (www.IntelSols.com), a well respected Business Intelligence and CRM consulting and education firm in 1992. Her company has successfully implemented over 150 Corporate Information Factory architectures in all industry areas.

About Embarcadero

Embarcadero Technologies, Inc. is a leading provider of award-winning tools for application developers and database professionals so they can design systems right, build them faster and run them better, regardless of their platform or programming language. Ninety of the Fortune 100 and an active community of more than three million users worldwide rely on Embarcadero products to increase productivity, reduce costs, simplify change management and compliance and accelerate innovation. The company's

flagship tools include: Embarcadero® Change Manager™, Embarcadero® RAD Studio, DBArtisan®, Delphi®, ER/Studio®, JBuilder® and Rapid SQL®. Founded in 1993, Embarcadero is headquartered in San Francisco, with offices located around the world. Embarcadero is online at www.embarcadero.com.