



Applied Data Resource Management

White Paper



Starting the  
Data Warehouse  
From a Data Model

Data Warehousing With Informix: Best Practices  
Prentice Hall ISBN 0-13-079622-0



A data model is a graphic representation of the data for a specific area of interest. That area of interest may be as broad as all the integrated data requirements of a complete business organization ("Enterprise Data Model") or as focused as a single business area or application.

Frequently a data model represents a functional business area (CUSTOMER, MARKETING, SALES, MANUFACTURING...) or a business area that is to be analyzed or automated (LEAD TRACKING, PROBLEM REPORTING, WARRANTIES...).

The function of the data model is to clearly convey data, data relationships, data attributes, data definitions and the business rules that govern data.

Data models are the accepted way of representing and designing databases.

The data models for a company tend not to change greatly over time unless the company changes the fundamental way that it does business. The way that data is used, the processes, can vary greatly even between organizations in the same industry.

However, the data required by companies within the same industry tends to be similar. This is why we can assume that data models have basic stability within the organization while process models are relatively unstable and constantly changing.

This common functional data is the basic premise that allows 'template' data models for an industry to be applied to individual companies operating in that industry.

Template data models are pre-built, fully-functional data models for a specific industry that closely approximate the results that would be achieved if you built them from scratch

Template data models can be built for every conceivable data modeling requirement, including the following types of applications:

- Enterprise
- Data Warehouse
- Business Area
- DSS/Data Mart
- Dimensional

Template data models of the highest standards have common characteristics:

- Constructed for a specific industry or industry segment
- Clear, unambiguous, detailed and fully attributed

Template data models are based upon detailed industry analysis that enables fully-attributed models to be developed. This attention to detail is what accelerates the planning, analysis and design phase and makes the use of template data models of real value.



Each entity should closely approximate a table that a designer would use to design an application and a DBA would use to build that application. Without identifying the columns that comprise a table, the job has just begun.

Every entity and attribute must be completely defined in industry terminology, including appropriate examples.

Example: [Split Dollar Life Insurance](#)

A life insurance policy in which premiums, ownership rights and death benefits are split between an employer and the employee, or between a child and parent.

The employer pays the part of each years premium that at least equals the increase in the cash value. The employee may pay the remainder of the premium or the employer may pay the entire premium. When the increase in cash value equals or exceeds the yearly premium, the employer pays the entire premium.

If the employee dies while in the service of the employer, a beneficiary chosen by the employee then receives the difference between the face value and the amount paid to the employer, whichever is greater.

During employment, the employee share of the death benefit decreases. If the employee leaves the employer, the employer has the option of surrendering the policy to the employee in exchange for premiums paid or selling the policy to the employee for the amount of its cash value.

There are two types of Split Dollar Life Insurance Policies:

(1) Split Dollar Endowment - in which the employer owns all policy privileges. The employees only rights are to choose beneficiaries and to select the manner in which the death benefit is paid.

(2) Split Dollar Collateral - in which the employee owns the policy. The employee pays premium contributions, which are viewed as a series of interest-free loans, equal to the yearly increase in the cash value of the policy.

The employee assigns the policy to the employer as collateral for these loans. When the employee dies, then the loans are paid from the face value of the policy. The beneficiary receives and remaining proceeds.



It is not unusual for a suite of industry template data models to be supported by 5-7,000 pages of bound documentation.

Every relationship must be named.

The data model should tell the story of the subject area of interest as you read the relationships between entities.

"... CELLULAR TELEPHONE operates under one RATE PLAN ..."

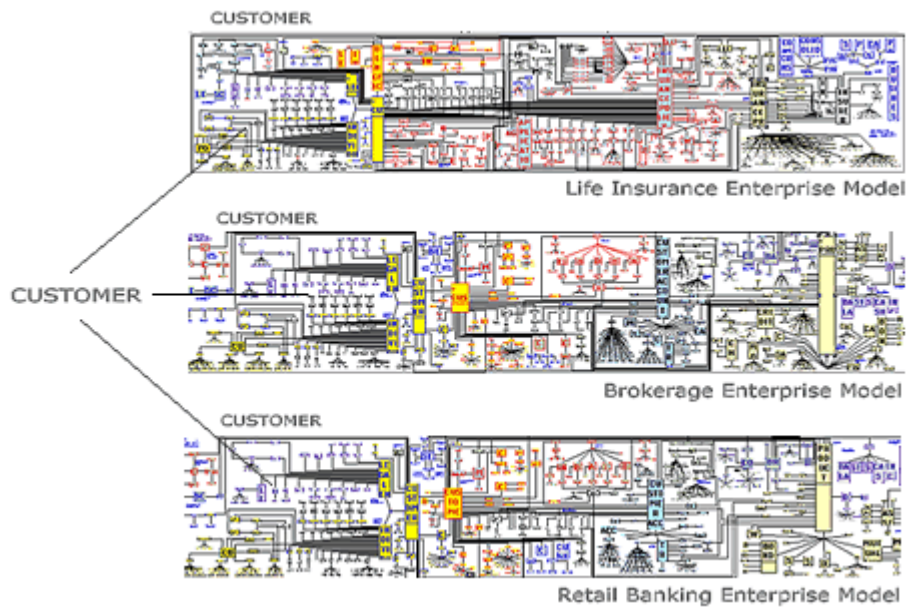
Data Models must be presented in large-format graphic representation, since few individuals read stacks of reports.

If the model is to be used, it must be presented in large-format graphics.

Template data models must be completely integrated.

This is the function of a strong CASE tool capable of supporting the development, integration and synchronization of models.

For example, as the definition of CUSTOMER is modified, it can immediately be propagated to all related models by linked schemas within the CASE tool environment.





### Why use 'template' data models?

The typical IS project consists of a long planning, analysis, design and implementation phase incorporating a host of hardware, software and staffing activities.

The most challenging, expensive and difficult to predict are the planning, analysis and design project phases. In fact, 60% of data warehouse implementations fail in these phases and not in the hardware or software selection process.

It is here that the decisions are made as to what data is important, how it will be structured and ultimately represented in the data warehouse and DSS and data marts.

There is no way to shorten the 12-18 months of planning, analysis and design. Nor is there a way to predict what will result, or if the organization will ultimately obtain a return on the investment, unless a set of detailed, industry-specific template data models are used to bootstrap these activities.

Template data models provide a close approximation of what would be achieved during that lengthy period at a small fraction of the cost. Their value must be immediately apparent upon inspection.

Template data models are not a 'plan-to-plan' or the starting point for some deliverable in the future. They are the deliverable that is intended to jump-start and dramatically shorten the planning, analysis and design phase.

It takes a few hours to determine their value.

Those few hours can save the project months of work and millions of dollars of costs. Template data models are flexible. They are designed to be modified, extended and integrated with other data models.

Applied Data Resource Management (ADRM) often presents an example of an Enterprise Data Model, which is approximately 20-feet in length and incorporates 350 tables and 2,500 columns, with 1,000 pages of documentation.

It is a compelling argument to tell senior management "You will never have less than what you see before you.

You can accurately predict project, resource and staffing costs by analyzing the areas of the models that need to be extended.

Furthermore, it can be installed and available in one hour."

Contrast this to an expensive, disruptive information engineering effort, which is frequently controlled by a third party, whose results will not be readily apparent for many months.



Template data models can be used immediately for:

- Building new data models
- Establishing a DA function
- Estimating resource requirements
- Gap analysis
- Industry knowledge transfer
- Project planning
- Quantifying ROI
- Standards definition
- Training and educating staff

Template data models can provide a clear, predictable view of the future without waiting.

### **The Cost of Entry**

If you look for template data models, you may be surprised by the fact that there are few to choose from and the level of detail can vary considerably.

Why is this?

To build good template data models requires a unique set of skills:

- Broad industry experience at each level
- Senior data modeling expertise
- Extensive DBA expertise
- Integrate the efforts of data modelers, architects and DBA's

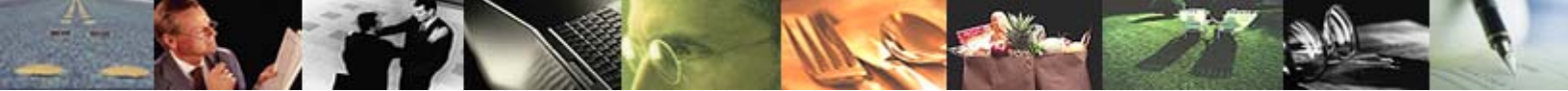
This combination of skills, expertise and long development cycles makes template data models unique. It also makes them extremely valuable.

### **What Industries are Supported?**

ADRM develops industry data models and packages them as a Data Environment for that industry.

Each Data Environment consists of an Enterprise Data Model, a multilevel Data Warehouse Model, 20-23 Business Area Models appropriately named, for example the "Semiconductor Data Environment".

ADRM supplies data models for 7 industries and 35 lines of business to companies throughout the world:



#### FINANCIAL SERVICES

Brokerage  
Commercial Banking  
Credit & Payment Card Services  
Credit Union  
Mutual Funds & Investment Management  
Mortgage Services  
Retail Banking

#### HIGH TECH PRODUCTS

Computers  
Medical Systems  
High-Tech Components  
Networking Equipment  
Semiconductor  
Software

#### INSURANCE

Automobile Insurance  
Healthcare Insurance  
Life Insurance

#### MANUFACTURED PRODUCTS

Agricultural Chemicals  
Automobile Products  
Aviation Products  
Consumer Packaged Goods  
Durable Medical Equipment  
Industrial Chemical  
Luxury Goods  
Pharmaceutical

#### RETAIL

Grocery  
Retail Products  
Restaurant Services

#### TELCO

Internet and Data Services  
LEC and Long Distance  
3G Wireless  
Wireless

#### TRANSPORTATION



Airline Industry  
Parcel and Mail Delivery  
Transportation and Shipping

These models are frequently combined or integrated with other models to meet the unique requirements of specific industries.

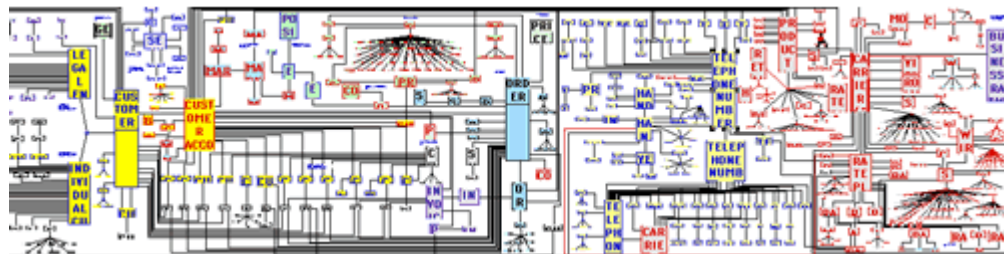
For example, the Banking industry is addressed by a single Banking Data Environment. However, it is frequently combined with the Insurance, Mutual Funds and Brokerage Data Environments to provide a comprehensive data model describing everything that a bank does in the current operating environment.

The data models are developed from a common core of building blocks so they can be rapidly integrated and still provide consistent definitions of Customer, Channel, Geography and other common data structures.

### **Everything Begins with the Enterprise Data Model**

Regardless of what is being built - data warehouse, DSS/data mart, application - it must be put into a larger context if it is to be successfully integrated with other systems.

That larger data context is the 'Enterprise Data Model'.



**Wireless Enterprise Model ("36x250")**

The Enterprise Data Model sets the stage by identifying all the data within the purview of a typical operating company in that industry - Customer Types, Channels, Beneficiaries, Life Insurance Products, Rate Plans, Business Ratios, SIC Categories and on and on.

Template Enterprise Data Models for specific industries are based upon in-depth analysis of that industry's common functions, business areas, terminology, data relationships, definitions, examples and business rules.

The Enterprise Data Model is the primary data tool for strategic planning, communicating data requirements throughout the organization, implementing integrated systems and organizing data in the data warehouse, DSS/data marts and applications.





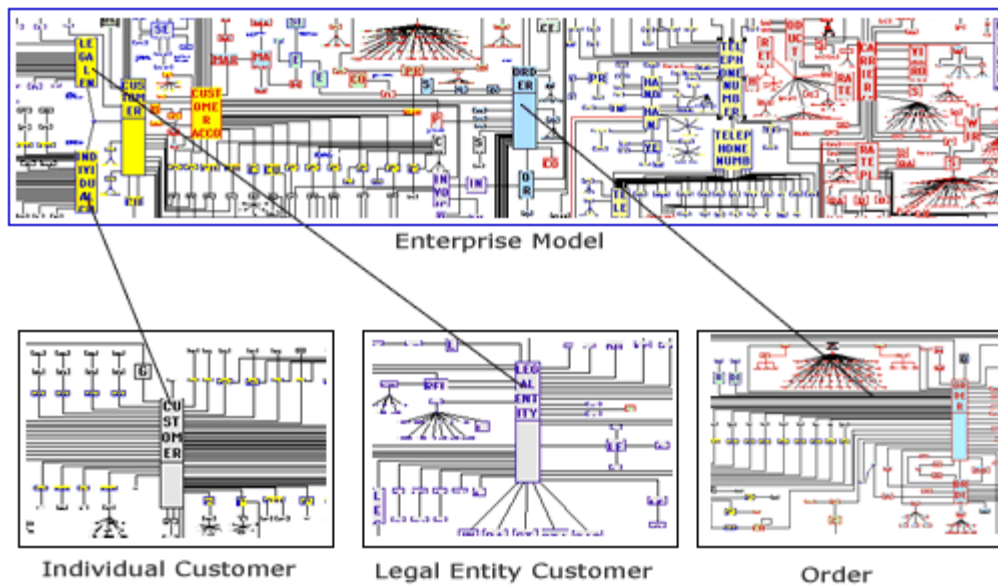
## Business Areas are the building blocks for the Data Warehouse

Developing the Data Warehouse does not start with that as the focus of analysis. It starts with the analysis of the Business Areas that contribute data to the Data Warehouse and receive information from the Data Warehouse. Business Area data Models describe lower levels of detailed data appropriate to building applications and DSS/Data Marts for a specific part of the business.

Each Business Area Model is constructed from a core of entities taken from subject areas in the Enterprise Data Model.

This insures that Business Area Models will be based on common key entities, have common keys, attributes and definitions through the data architecture.

This approach also supports consistent integration of existing data models and development of new data models.



Business Area Models contain the greatest level of detail and represent the lowest level of data granularity in the data model hierarchy.

As new information is learned about business areas, that information is added to the corresponding Business Area Model. That information may also be incorporated into the Enterprise Data Model, Data Warehouse Model or application data models.

Using Business Area Models as a basis for analysis and design provides a solid foundation of industry-specific knowledge that will accelerate planning and development.



To understand the interrelated requirements of the Data Warehouse requires an understanding of a host of seemingly unrelated data.

Examples of Business Area Data Models are:

- BOOKINGS/BILLINGS/BACKLOG
- CHANNEL
- COMMISSIONS
- CONTRACT
- CUSTOMER
- CUSTOMER SALES
- FINANCIAL
- FORECAST
- GEOGRAPHY
- HR/EMPLOYEE
- INVENTORY
- MARKET
- MARKETING EVENT
- ORDER
- PRICING
- PROBLEM REPORTING
- PROSPECTIVE CUSTOMER
- MANUFACTURING/SHOP FLOOR CONTROL
- TERMS & CONDITIONS
- TRAINING & EDUCATION

Individual models can easily be combined or integrated to create other models.

For example, a "Customer Care" prototype data model may be quickly built from 7-8 individual Business Area Models in a few hours.

#### **Using "template" data models to Build the Data Warehouse**

Once the Enterprise Data Model and Business Area Model components are in place, it is possible to introduce the template Data Warehouse Models.

The Data Warehouse Model represents the integrated decision support and information reporting requirements of the business.

The Data Warehouse is the center of the decision support and reporting data architecture, the ultimate source of clean, consistent data for the entire organization.



The Data Warehouse may be surrounded by any number of functional decision support systems or "data marts" serving the associated functional business areas. As data moves from the Data Warehouse to local DSS or "data mart" systems, control of the data is turned over to local administrators. The Data Warehouse remains the consistent source of reliable data.

The Data Warehouse Model is constructed directly from the same-industry Enterprise Data Model, which is the high-level data blueprint describing the organization's integrated information requirements. This insures that the collective information requirements of the enterprise are represented in the Data Warehouse Model.

Each subject area in the Enterprise Data Model has a corresponding Data Warehouse Model component.

ADRM implements the Data Warehouse Model in two distinct functional levels:

- Decision Support Data
- Summarized Data

The Level 1 - DSS data model describes data at the lowest level of detail appropriate for detailed analysis and decision making:

- "Orders from a specific Customer for a period of time"
- "Quantity sales for a specific Channel for a period of time"
- "Sales revenue by Product via a Channel during a period"

The Summary model depicts summarized data based upon data defined in the DSS model.

This is the level of data that senior management uses to base their decisions:

- "Total sales revenues from Orders for a period"
- "Total New vs. Rebuy Orders for a period from all Customers"
- "Total quantity sales from all Channels for a period of time"
- "Total sales revenues from sales of a Product for a period"
- "Total sales revenues from sales of a Product Line"
- "Total sales revenues from sales of all Products for a period"
- "Period Debt To Equity Ratio"

It might seem difficult to develop a 'template' Data Warehouse Model for a company. In practice it is an exercise primarily in decision making and secondarily in model building.

- "What are our Customer Types?"
- "What are our Markets and Market Segments?"
- "How will our Geographies be defined?"
- "How will we define Product Families and Product Lines?"



A building block approach can be applied to constructing the data warehouse.

- The Enterprise Data Model sets the stage.
- Business Area Models develop the detail from the Enterprise Data Model.
- The Data Warehouse Models are first-cut, good practices data models containing a view of what an well-rounded data warehouse might contain.
- Detail data from the Business Area Models are added to the warehouse data model in a series of building block steps.
- Legacy data is considered and integrated into the Data Warehouse Model as needed. The Data Warehouse is built from existing template data model components representing the way the business intends to do business, which are then modified to meet the realities of legacy data and existing applications.

The process is greatly accelerated by utilizes standard industry data building blocks that can readily be modified, extended or integrated to meet specific data requirements.

Each of these models are modified and contribute data to related models. Because the models are broken up in functional building blocks, progress can be made in many areas without waiting for any single decision to be made. Product data can be designed without waiting for the final Customer data structures. Channel values can be defined later while the foreign key relationships are defined immediately.

Progress can be made in many areas confident that the final results will fit together.

The level of detail of the Business Area Models is consistent with that of the DSS/data marts. This makes it relatively easy to develop DSS/data marts that dovetail with the Data Warehouse. It also makes it simple to promote data structures from the Business Area Models into the Data Warehouse Model or related Business Area Models.

A good example is Inventory. The ADRM industry Enterprise Data Models do not describe inventory in detail because the models would become too dense. That detail is carried in the Inventory Business Area Model.

However, if inventory data is of enough significance to the business organization, it can easily be integrated at any level of detail from the Inventory Business Area Model into the Data Warehouse Model.

What would have been a significant analytic exercise can be dramatically accelerated by starting with detailed good baseline models that can be quickly modified to meet specific requirements.

The focus of the effort becomes decision-making not model building.



### **Comparing the Costs**

A typical information engineering project is measured in many millions of dollars.

A target return on cost of models should be a minimum of 20:1. A project should easily see \$1-2 million dollars in benefits by introduction of the template data models. Even if the template data models are used as an advanced starting point from which to focus the project, they will have repaid their value many times over.

What cannot be quantified are the intangibles of using template data models:

- Directing resources upon focused objectives immediately
- Reduced disruption to the organization
- The ability to clearly convey objectives at a detailed level
- Identifying clear roles and objectives for participants

It is always wise to consider the use of template data models.

### **Guidelines for Using Template Data Models**

- Clearly define your data objectives.
- Are template data models appropriate for your objectives?
- Is the level of detail appropriate?
- How much additional work will be required?
- Is the terminology consistent with that of your company?
- Does your staff see the value immediately?
- Develop an Enterprise Data Model either as a primary activity or as an adjunct ongoing activity.
- Use template data models as a tool to make your job easier, educate your staff and help you better understand what can be achieved.
- Don't consider template data models to be turnkey solutions.

Understand their strengths and weaknesses relative to your objectives.

They one more tool to be utilized to address specific objectives.



- Look for a conservative return of 20:1 on the price of template data models.
- Strive to simplify the data required by your business at every opportunity.

This will dramatically streamline IS systems with corresponding savings. It will make integration of systems easier. It enables the staff to understand the workings of the company and contribute to a broad range of activities.

- **Use true experts.**

Assigning someone to CUSTOMER that has no experience with that subject area will not make it better. Consider assigning a data modeler and an acknowledged expert to address the most critical areas of the project.

- **Hire the best people available.**

Quality people are needed to produce quality results. Interview candidates in depth. They should be able to demonstrate how they would tackle problems and present a clear picture of the ultimate results to be achieved.

Beware of concepts and terminology that you cannot understand. You are planning systems for your business. Make sure the results will be acceptable to your business community.

- **Never assume that you can't do a good job yourself.**

If you have the time and resources, your staff are the best people to build your data models. Why? Commitment. You will live with the results. It is your company that you are changing for the future. No one else can possibly have that same level of commitment and ownership that is so critical to success.

- **Don't be afraid to succeed.**

Do not agonize over making decisions. Make good, solid decisions. If it is reasonable, well-thought-out and consistent with how you intend to do business, consider it done. Move on to the next challenge.

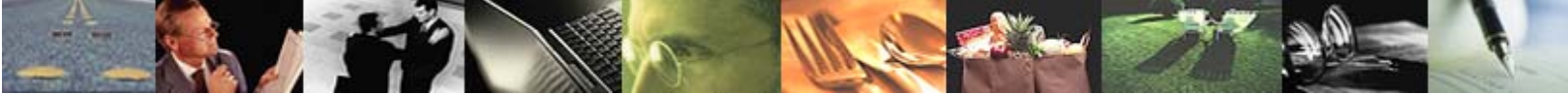
Common sense is an underestimated business skill.

- **If you make a mistake - fix it.**

It is only natural to learn new things as the project and your skills evolve. If you have made a serious error in defining PRODUCT, fix it now.

The company will thank you and your commitment to the job will be acknowledged.

It is a fascinating, rare opportunity to look to the future and change the way a company operates.



Enjoy the camaraderie that develops as you work with the best and brightest people in the company. Celebrate your accomplishments at every opportunity.

### **Summary**

To benefit from good data technology, a business must make the commitment to make their data good. Only then can they apply a broad spectrum of tools and technologies consistently across the business.

Template data models provide a data platform for designing new data and integrating legacy data at an accelerated pace with reduced overall costs.

The 60% failure rate in building data warehouses is a function of planning, analysis and design - and not due to the hardware or software selection process.

Template data models are one of the few tools that can be applied immediately to produce immediate results at nominal costs.

The cost to investigate template data models is usually a few hours of investigation.

The results may change the way your company does business far into the future and make your job much easier.

### **About the Author**

[Larry Heinrich](#) is founder and president of Applied Data Resource Management. He holds various degrees in Mathematics and Computer Science and has 30 years of industry experience.

He can be contacted at [heinrich@adrm.com](mailto:heinrich@adrm.com).

### **About Applied Data Resource Management**

Applied Data Resource Management, Inc. specializes in clearly defining the information requirements of world-class organizations in a variety of industries and architecting intellectual property based products to help organizations in those industries more effectively capitalize upon their information assets and opportunities.

More information about ADRM can be found on the web at [www.adrm.com](http://www.adrm.com) or contact Kevin Schofield, VP Sales and Marketing, at [schofield@adrm.com](mailto:schofield@adrm.com).