



claudia  
imhoff

Reprinted from  
**DMReview**

# Lessons from the Farm – Managing the Data Delivery Process

*Claudia wishes to thank Frank Cullen from Blackstone and Cullen Consulting for his insightful input into this column.*

**M**ost farmers are very organized. They manage their environments with great care and exacting precision. A field is tilled and earmarked for a specific crop. The seeds for that crop are planted at precise intervals along the rows of dirt; the crops are harvested and segregated so they aren't mixed up. This maximizes the level of quality and efficiency – both mandatory features for a farm to survive today.

A farmer would never even think of mixing his seeds together in a bag and simply winging them out into the freshly plowed field. Chaos would result; his very livelihood would be in jeopardy. He would end up with a stock of corn growing next to alfalfa, which would be growing next to wheat, etc. Harvesting would become a monumental task in which different plants would be harvested at different times by hand! I think there is a lesson here that we can learn from these well-organized and competent people.

Think of the Corporate Information Factory (CIF) as a well run farm. Those of you familiar with this architecture (Figure 1) are aware that it is easily split into two halves (or sets of fields, if you will) – each with its respective large process and data stores.

One half deals with “getting data in” and consists of the opera-

tional systems, the data warehouse and/or operational data store, and the complex process of data acquisition. Much has been written about these components, especially the extract, transform and load (ETL) part of data acquisition. The ultimate deliverable for this part of the CIF is a repository of integrated, enterprise-wide data for either strategic (data warehouse) or tactical (operational data store) decision making.

The other half of the CIF deserves some more attention. It is summarized as “getting information out” and consists of the data delivery process, the variety of marts (data and oper) available for the business community’s usage and the decision support interface (DSI) or technologies that access the marts and perform the various analytics or reporting needed by the business community. The ultimate deliverable for this half of the

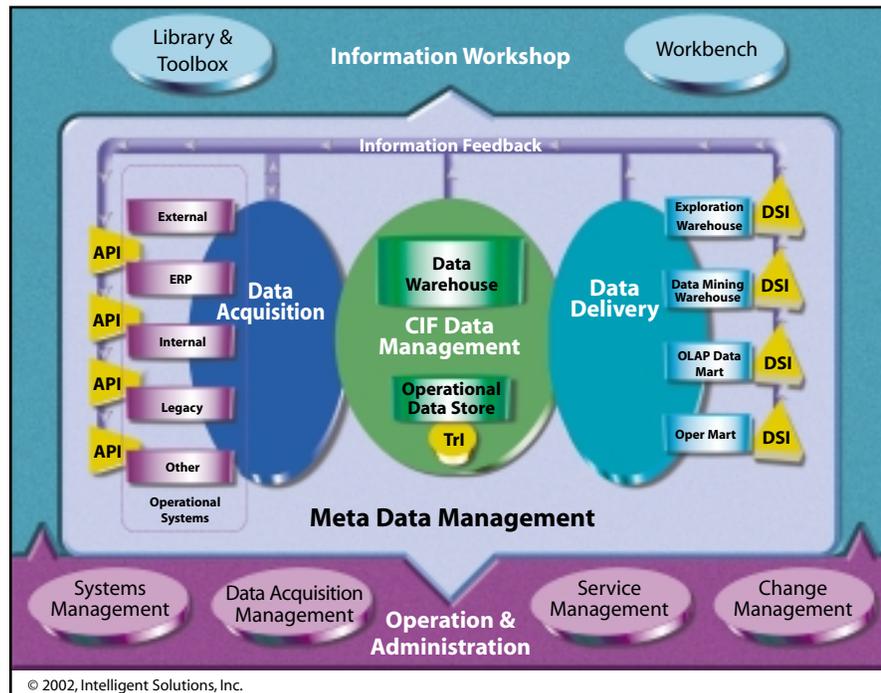


Figure 1: The Corporate Information Factory

CIF is an easily used and understood environment in which to perform analyses and make decisions. Most of the highly touted business intelligence (BI) benefits are derived from getting information out – data consistency, accessibility to critical data, improved decision making, etc.

However, is this really being achieved? I don't think so. Unfortunately, many corporations have not followed the architecture as closely as they should have. They have created an environment that is all too similar to a farmer putting all his seeds into a bag and blasting them out – willy-nilly – into his fields. Let's look at what we have created in more detail.

The construction of the data warehouse is now well documented and has eliminated much of the chaos in terms of getting consistent data from our operational systems. We are now able to clean up the data as well, improving its quality significantly. We place this data in easily accessed database technologies with the idea that data marts can be quickly built from this resource.

Unfortunately, we have not paid as much attention to the creation of marts as perhaps we should have. With the warehouse in place, it becomes very easy to create cube after cube, star schema after star schema, data set after data set, seed after seed, from this repository – but with minimal management or control over these constructs. Redundancy and inconsistency have crept into this half of the architecture, significantly threatening the promised benefits. Figure 2 shows what is happening here.

Many companies have more than one ETL tool used for the delivery of data

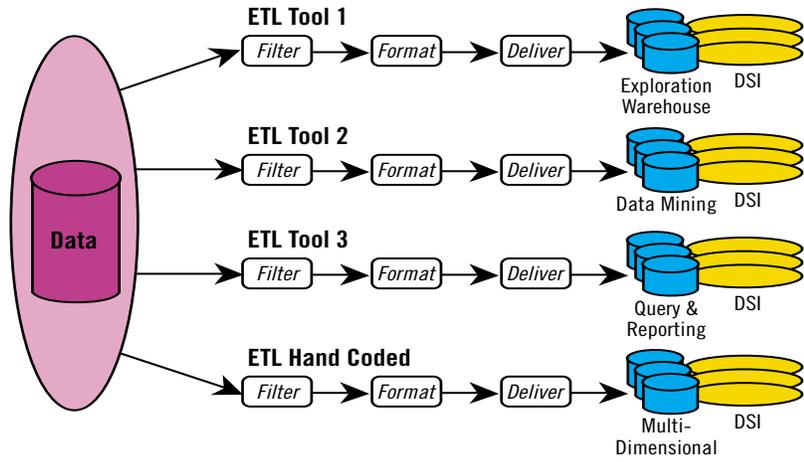


Figure 2: Chaos in Data Delivery

into the marts. Some are used to create the data warehouse and marts; others come with the DSI tool of choice or even the data mart database. We have also used hand-coded data delivery programs – all now run rampant through the warehouse extracting data and winging it out to marts at will. This by itself would not be a problem if it were a managed process. The difficulty is that the discipline in many organizations is not in place to ensure that these processes are efficient and administered. What I see more often today are the following situations:

- Duplicate marts being created – each with virtually identical functionality but under the control of different groups within the business. There are many BI functions that are needed in more than one department, by more than one group of users. Inefficiency and non-productivity are the result of duplication.

- Inconsistency in terms of the extraction timing for data delivery, leading to numbers that don't match even though the functionalities appear similar. For example, a product profitability mart may be built for finance and another product profitability mart built for sales. Although the functionality may be identical, one is created on a daily basis and the other on a monthly basis. There is no reason to expect these two to generate identical profitability numbers.
- Marts that are no longer being used. Because there is no management process in place, many companies are creating marts that are no longer needed by the business. The proliferation, especially of cube technology, has been a significant contributor to this unfortunate problem. It is perceived as easy to create a cube and then simply continue to recreate it day after day, month after month, year after year, whether someone is using it or not. This is a terrible waste of resources!

The resulting situation is not pretty: multiple tools means multiple skills required, reusability of data delivery code may be constrained or limited, meta data becomes encapsulated within various tools and is not sharable across tools, inconsistency of the data used is highly probable and the overall environment is more costly to maintain and sustain.

What is needed is a new paradigm, a return to the principles of the architecture, a shift in our thinking about data delivery. It cannot be an unmanaged, uncoordinated set of processes as represented in

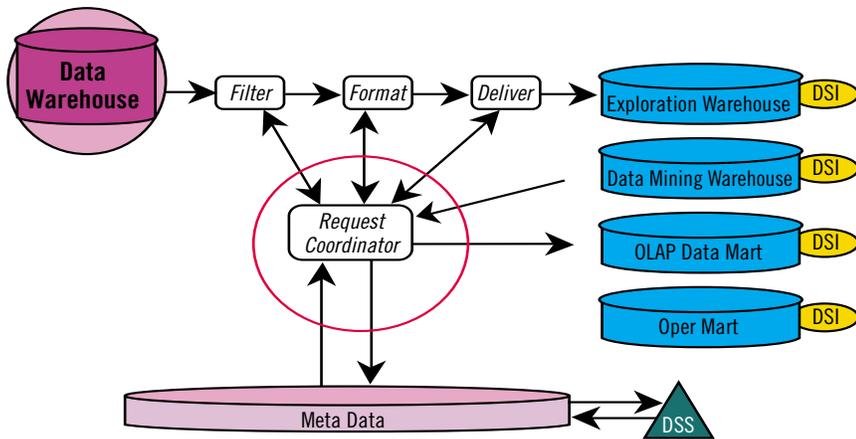


Figure 3: Data Delivery with a Request Coordinator Function

---

Figure 2. We must create a consistent, documented and managed environment that starts with a request coordinator process. Figure 3 demonstrates a managed function in the CIF.

The request coordinator is like the farmer who plans his next season carefully, determining what seeds will be planted, which fields will have what crops, where efficiency of scale, market value, and time to market (harvest schedule) play a role, etc.

In the CIF, the request coordinator first captures the business user requests, prioritizes them and then profiles them to fully understand the request. Meta data plays an important part in this step – it is used to determine whether a new mart is warranted or an existing one can be enhanced to accommodate the request. If a data mart that can satisfy the request already exists, then the function simply

gives the users access to the mart, perhaps adding a bit of new data, a new report or creating a view specifically for that set of users. If a mart does not exist, then the coordinator must begin the process of filtering the right data from the warehouse, formatting it to the correct technological format, and delivering that data to the new mart per the requested schedule.

In researching this column, I looked at a number of technologies that could help with this data delivery management problem. Certainly it is possible for you to use your existing ETL tools or even the many bulk data movement technologies (IBM, Microsoft, iWay and other EII capabilities). However, you still need to create the request coordinator function and manage the meta data associated with the data delivery process.

I also found a new technology

offered by Certive that manages not only the creation of marts but also the meta data and business rules for each mart creation. This new technology is a bright spot in our industry and deserves consideration.

In any case, regaining control over the data delivery process requires a shift in your existing architecture; however, the benefits of this shift outweigh any disruption to “business as usual.” These include reusable delivery code, managed meta data and business rules, potential usage of virtual marts and lowered overall data delivery costs. 

---

Claudia Imhoff, Ph.D., is the president and founder of Intelligent Solutions ([www.intelsols.com](http://www.intelsols.com)), a leading consultancy on CRM and business intelligence technologies and strategies. She is a popular speaker and internationally recognized expert and serves as an advisor to many corporations, universities and leading technology companies. She has coauthored five books and more than 50 articles on these topics. She may be reached at [cimhoff@intelsols.com](mailto:cimhoff@intelsols.com).