

DMReview

corporate information factory

CIF^e: Evolving with the Times

By Jonathan G. Geiger

Gasoline prices have reached about \$3.00 per gallon, and in the aftermath of the recent hurricanes, there is talk of much higher prices. The increasing prices have impacted the technologies that are being incorporated into cars as well as our selection and use of automobiles. The manufacturers have been emphasizing fuel efficiency for several years and are introducing new technologies, including more streamlined designs, electric engines and hybrid engines to address the rising fuel costs. We as consumers are moving away from SUVs and are changing our driving habits.

Corporate Information Factory Evolution

In the last few years, we've seen changes in the business intelligence (BI) world that are analogous to the changes impacting our driving habits. Companies are applying their BI environments to support customer relationship management (CRM) and business





process management (BPM) applications in a near real-time manner; they are increasingly incorporating the results of analytic applications into their operational environments and the BI environment is frequently used to satisfy regulatory requirements. Just as the automobile manufacturers are reacting to the changing demands, technological advances described later in this article have been introduced. The Corporate Information Factory (CIF), which provides a resilient framework for BI applications, also needs to evolve. The evolution shown in Figure 1 is largely the result of a collaborative effort by Claudia Imhoff, Colin White and myself. We call it the CIF^e, or the Extended Corporate Information Factory.

At first glance, you may think that this is a complete departure from the old CIF. However, a closer examination will show you that we've done the equivalent of providing a hybrid engine while maintaining the basic functionality and principles of the CIF from its inception. These principles include:

- The program orientation of the BI environment,
- The sourcing of data from the operational systems environment and external data stores,
- The storage of an enterprise view of the data in a data warehouse and operational data store (ODS),
- The delivery of that data to the business community through data marts that are tailored to the business users and their applications or directly from the ODS, and
- The inclusion of meta data management throughout the environment.

The new depiction of the environment emphasizes a few other principles. While all of these were supported within the CIF framework, their inclusion was not obvious in the pictorial representation. The new CIF^e icon enhances the visibility of:

- Options for data flow using available technologies such as enterprise information integration (EII) and enterprise application integration (EAI),
- The meta data repository itself (in addition to the meta data management function),
- Additional components of environment management such as governance and centers of excellence, and
- Knowledge management as an integrating point for providing and using

information.

Let's examine each of these areas individually.

Data Flow

The traditional CIF approach included a data acquisition process to bring data into the data warehouse (and/or operational data store) and a data delivery process to move that data into the data marts. To understand how this is still supported within the CIF^e, we need to look deeper into the data integration and data delivery process. As we look into this process, we see three techniques for integrating data within the environment, as shown in Figure 2, with the data consolidation option being the traditional ETL approach.

Data Consolidation. With data consolidation, data is physically extracted, transformed, cleansed and integrated, and loaded into a downstream data structure. For data acquisition, the data is extracted from the operational sources, integrated, cleaned up and finally transformed into the corporate standard and loaded into the data warehouse or ODS; for data delivery, the data is extracted from the data warehouse or ODS, formatted for a particular application and then loaded into the marts. This process remains the preferred approach for populating a data warehouse supporting strategic analysis and is shown in Figure 3.

Data Propagation. The previous version of the CIF included an information feedback loop. One of the ways data propagation can be used is to feed the results of analytic applications to upstream operational systems or the ODS. For example, analytic applications may use the integrated view of customers to identify angel and demon customers.¹ Angel customers are the most profitable customers, and these deserve special treatment to maximize the potential for retaining them. Demon customers are the customers who cost the company money, and special actions may be appropriate for these customers, as well, to either increase their profitability or influence them to go elsewhere. By providing the operational systems that contain the results of customer analyses, the customer-facing employees will have the information at their fingertips and will be able to adjust their activities based on the value of the customer to the firm. Data propagation is supported mostly by EAI technologies.

Data propagation is used for mass movement of data. However, it should be noted that this technique is not appropriate when complex and complicated integration or transformation is required before moving the data.

Data Federation.

Data federation, which is supported by EII technologies, provides a consolidated view of data to support time-sensitive applications such as business performance management and executive dashboards. An example of data federation is to provide a view of data in the source system through a virtual operational data store or data mart for use by the business community. The data is not actually moved into an intermediary data store; rather, the user is provided a “virtual” view of the compiled information. Data federation is very useful for integrating similar information from multiple sources. For example, a company may have multiple systems which capture product sales at different locations. Data federation is an appropriate way to provide the total sales information instantly.

Data federation has some very valuable applications, but it is not a solution for all data integration needs. If a persistent data store is needed (e.g., data warehouse), this is not the appropriate approach.

Meta Data

Meta data has its roots in the data dictionaries that we used in the 1960s and 1970s to capture technical information about our application systems. Within the BI environment, technical meta data is expanded to include not just information about the data storage (e.g., what sources were used, what targets were used). It also includes information about the data integration and movement (e.g., transformation rules, data quality processes and audit information), as well.

Within the traditional application system environment, access to data was controlled largely through programmed transactions and predefined reports. As such, procedure manuals often included appropriate explanatory information. With

the advent of BI, users gained increasing flexibility in the way they accessed information. To ensure that they understood the information they obtained, business meta data entered the landscape. With business meta data, the users were able to get business definitions, business rules and other business-clarifying information for the data elements they were accessing.

As the usefulness of good meta data became widely recognized, a third type of meta data arose. Administrative meta data was designed to help the people who operate, maintain and support the BI environment. Administrative meta data provides information concerning the performance of the environment, the quality of the data, data that is commonly used, who is using the data marts, and so forth, so that actions can be taken proactively and intelligently to ensure a smooth, cost-effective operation.

During the last few years, companies have increasingly applied their BI environments to support regulatory compliance. For example, to satisfy certain requirements of Sarbanes-Oxley legislation, companies must be able to quickly substantiate the transparency and validity of their data. Meta data is created to meet this need and, in doing so, the meta

data repository becomes a valuable information source in-and-of itself. Within the CIF^e, we have increased the visibility of the meta data repository by showing it as an explicit repository in the environment.

Environment Management

The outer band of the CIF^e shows the major components of the environment management function. These components provide a more modern approach to grouping major activities that must be performed to ensure that the BI environment operates smoothly and cost-effectively, as well as increases in value to the organization as the business learns to leverage and expand its application. A quick description of each of the six major sets of activities within environment management follows. We will expand on these in future columns.

- *Governance*: People and processes for controlling and coordinating the environment and the individual BI projects;
- *Infrastructure Management*: People, processes and technologies for ensuring that the environment operates smoothly and reliably;
- *Center of Excellence*: People, processes and technologies for promoting collaboration

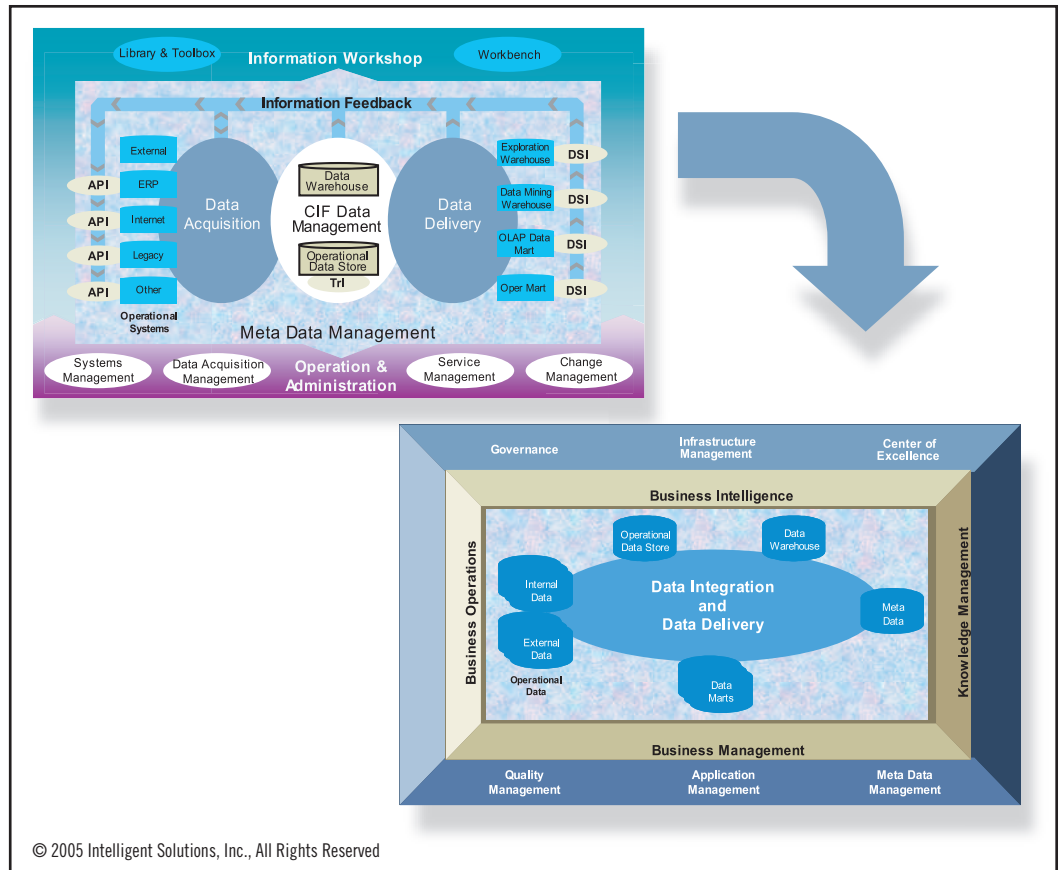


Figure 1: Corporate Information Factory Evolution

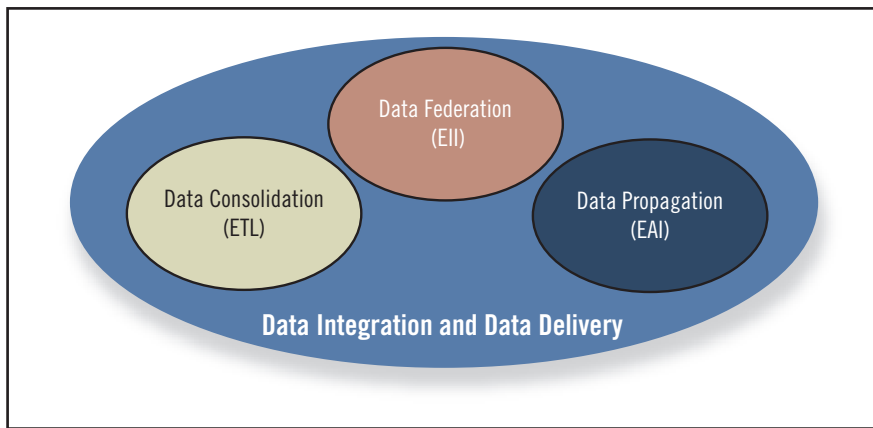


Figure 2: Data Integration and Data Delivery

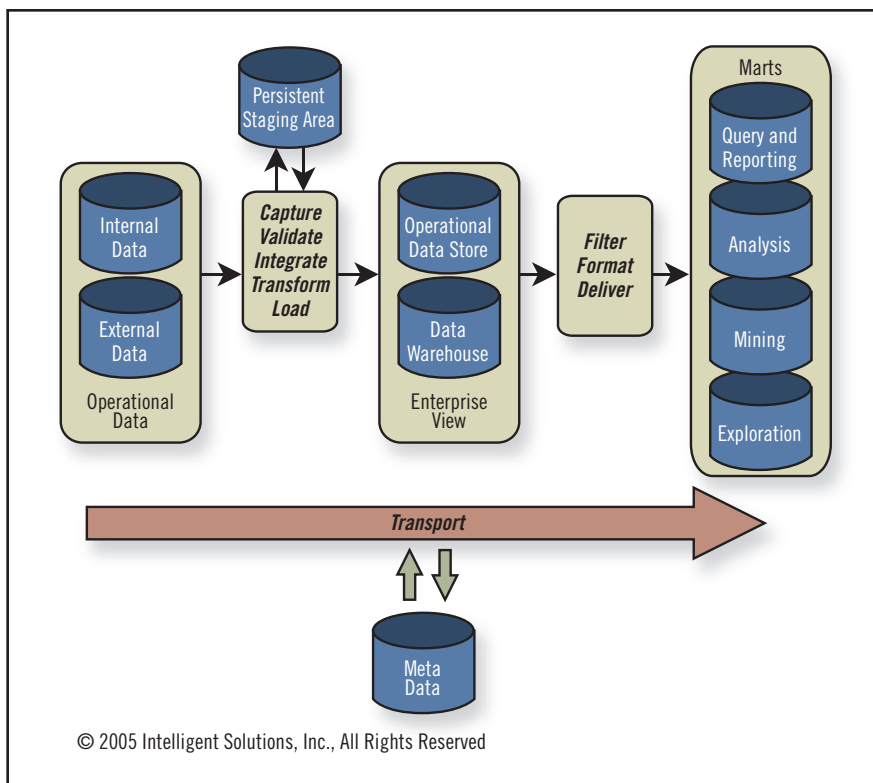


Figure 3: Traditional ETL

- and the application of best practices;
- **Quality Management:** People, processes and technologies for ensuring that the data quality meets business expectations;
 - **Application Management:** People, processes and technologies that enable application of the information to provide business value; and
 - **Meta Data Management:** People, processes, technologies and data stores for managing the information about the enterprise's data resources and activities.

Knowledge Management

We've described the interior of the CIF[®], as well as its outer layer. There is a subtle


change in the middle layer, as well. The middle layer includes three components that were explicitly addressed in the previous versions of the CIF - business operations, business intelligence and business management.

- **Business Operations** is the set of day-to-day activities of the business. It is within this function that we find the operational transaction processing systems and external data. Typically, the processes that support this function are fairly static and change in quantum leaps.
- **Business Intelligence** is the application of integrated strategic data to understand where the enterprise and its market-

place have been, predict where they are going and determine appropriate changes to solidify or change that direction. BI applies information from the data warehouse and associated marts. The processes and applications, which often change and evolve, include customer retention analysis, customer and product profitability analysis and supply chain analysis.

- **Business Management** is the application of integrated tactical data to understand and manage the performance of the enterprise. Typically, business management uses information brought together through EII or information from the operational data store for applications such as CRM and BPM as it carries out its enterprise strategies.

Knowledge management brings the three functions together. It applies enterprise portal, collaboration, content management, BI and process management technologies to enable users to locate, share and communicate about the business content (data, information, expertise and applications) they need to do their jobs.

In an August 2004 survey, more than 40 percent of the companies surveyed indicated that they were using an architecture that included a central store of enterprise data feeding data marts for user access. The CIF is the best-known architectural framework promoting that approach. Since it was first created, business needs and new technologies constantly evolve, and the CIF must evolve to keep up with those changes. CIF[®] is the most recent of these evolutionary changes. The latest set of adjustments addresses changes in the business environment and incorporates new technologies and techniques while preserving the principles that underlie the framework. It provides a well-paved highway on which businesses can drive into the future in a safe and resource-efficient manner. 

References:

1. Selden, Larry and Geoffrey Colvin. *Angry Customers and Demon Customers*. New York: Portfolio, 2003.

Jonathan G. Geiger, executive vice president at Intelligent Solutions, Inc., has been involved in many Corporate Information Factory and customer relationship management projects within the utility, telecommunications, manufacturing, education, chemical, financial and retail industries. He has authored or co-authored numerous articles and three books, presents frequently at national and international conferences, and teaches several public seminars. Geiger may be reached at JGeiger@IntelSols.com.