

The BI Watch
Who's On First
Richard Hackathorn
(Submitted for the March 2004 Issue of DM Review)

A classic comedy routine was *Who's On First* by Abbott and Costello, performed many times around 1950.¹ The humor comes when the listener realizes that the actors are using the same words but with entirely different meanings. The noise level ever increases and the frustration mounts as Abbott and Costello struggle fruitlessly to achieve a coherent conversation.

Unfortunately I have the same feeling about a defining issue in Business Intelligence (BI) and Data Warehousing (DW). The goal of DW and now Enterprise Information Integration (EII)² is a single version of business reality by... consolidating content so that business processes appear consistently across disparate information sources. On the other hand, the goal of Enterprise Application Integration (EAI) is a single version of business reality by... consolidating functionality so that business processes interact seamlessly across disparate application systems.

So, who's on first? EII and EAI are using the same words but with different meanings. If your budget is not being squeezed, enjoy the comedy of the situation. Most cannot!

The problem is that corporations treat EAI and EII as different approaches with different goals supported by different groups who compete for the same resources. Instead, EII and EAI should be considered as two sides of the same coin.

Let's explore this (comical?) situation in Figure 1, which shows a simplified architecture incorporating both EAI and EII.

>>> Insert Figure 1 <<<

Various application systems performed their specific functions. In addition, they feed application events onto the EAI bus and feed application data via ETL procedures into the data warehouse. For example, the EAI bus may see...

commit-sales (time,customer,product,quality,salesperson...)

while the data warehouse may see...

sale (sale-id, time,customer,product,quality,salesperson...)

So, what is Frank doing, as compared to Sandy?

¹ For background on Abbott and Costello, see <http://www.abbottandcostello.net/>. The script for the *Who's On First* routine is at <http://www.abbottandcostello.net/who.htm>. An excellent rendition of the original routine was done by Larry & Timothy Cappetto at http://directmediapro.com/abbott_and_costello_who.htm (toward the bottom).

² In this article, EII refers to the next step in the evolution of the Data Warehousing architecture, merging the Data Warehouse that persists historical snapshots with federated query that taps into reservoirs/streams of disparate data.

Frank is using Business Activity Monitoring (BAM) tools to analyze business events as they happen. He is focused on the day-to-day running of the business, such as monitoring hourly sales levels through the day as compared with expected levels at that time on previous days. Low sales levels may motivate Frank to correct operational problems in store staffing and system disruptions.

On the other hand, Sandy is using BI tools to analyze business data historically. She is focused on planning, such as trends in daily sales levels as compared with previous months. Low sales levels may motivate Sandy to correct tactical problems in marketing messages and store promotions.

Diaz Nesamoney of Celequest suggested an analogy: “Frank is driving to an appointment using a GPS to direct him, while Sandy is planning her vacation using a tourist guidebook.” The implication is that their roles are cleanly separated by different analysis horizons – short-term (hourly variations) for Frank and long-term (monthly variations) for Sandy.

This separation of analysis horizons is valid if business requirements are stable and predictable. However, all businesses globally are experiencing increasing turbulence in their marketplace and technology, forcing them to think and act short-term and long-term concurrently. For instance, a catastrophic event may happen at 8:46 one morning that negates any trends predicted by historical data.

To compensate for the turbulence, Frank is improving his short-term analyses by adding the context of a historical store for business activity. And, Sandy is improving her long-term analyses by adding the context of real-time feeds into an Operational Data Store (ODS). From the previous analogy, Frank better have a good map, and Sandy better have an up-to-date guidebook.

Unfortunately, Frank and Sandy often do not collaborate with one another, being at opposite corners of the enterprise architecture. When they do talk, they use the same words but with different meanings. In fact, they are probably in different IT groups competing for the same IT funds, resulting in a power struggle for architectural control.

Competition is good to a certain extent, but this is not healthy for the enterprise in the long term. Colin White of BI Research remarked, “As a company pursues real-time warehousing, there is usually political divisions among the groups involved, resulting in a problem of data disintegration.” Likewise, there is the struggle for market share among emerging vendors who play at the boundaries of EII and EAI.

So, what is the solution? Barry Devlin of IBM suggests an intermediate solution using a federated query approach, such as found in IBM DB2 Information Integrator. Figure 2 shows the federated query layer tapping into both event and data stores to provide a more unified view. Replacing Frank and Sandy, Carol is now enabled to analyze both the event stream in the context of historic warehouse data and the data trends in the context of real-time business activity.

>>> Insert Figure 2 <<<

This intermediate solution helps but conceals two important problems.

First, there is significant redundancy in the architecture of this intermediate solution. Flows and stores should be consolidated to reduce the number of components required.

Second, there are deep design issues pleading for a unifying theory of EII and EAI. What is the proper latency for capturing business events? Likewise, what is the proper persistence for storing business data? And, many more questions...

We need a Codd-like theory that guides us to minimize the functional dependencies among process events and data objects...or something like that. The new goal should be a unified view of business reality, just as the goal of the data warehouse was a single version of business reality.

This unifying theory will not come easily since it is closely intertwined with that dirty word 'metadata'. "There is a lot of confusion out there, and semantics is part of the problem. It is ironic that we're having a semantic problem that is all about metadata, which is all about semantics," as Darren Cunningham of Business Objects ponders.

Data integrity must also be addressed by this unifying theory. As banking and other industries are moving to Straight-Through Processing, the luxury of consistency checking and data cleansing during batch updates vanishes, as observed by Devlin.

It will take several years to sort through these tough problems. Meanwhile, who's on first at your company?

Dr. Richard Hackathorn is president and founder of Bolder Technology, Inc., a twelve-year old consultancy in Boulder, Colorado. Richard has over 30 years of experience in the IT industry and is a well-known technology innovator and international educator, conducting professional seminars in 18 countries. He has written three textbooks entitled Enterprise Database Connectivity, Using the Data Warehouse (with W.H. Inmon), and Web Farming for the Data Warehouse. He earned his B.S. degree from the California Institute of Technology and his M.S. and Ph.D. degrees from the University of California, Irvine. To contact, send email to richardh@bolder.com.

1,134 words

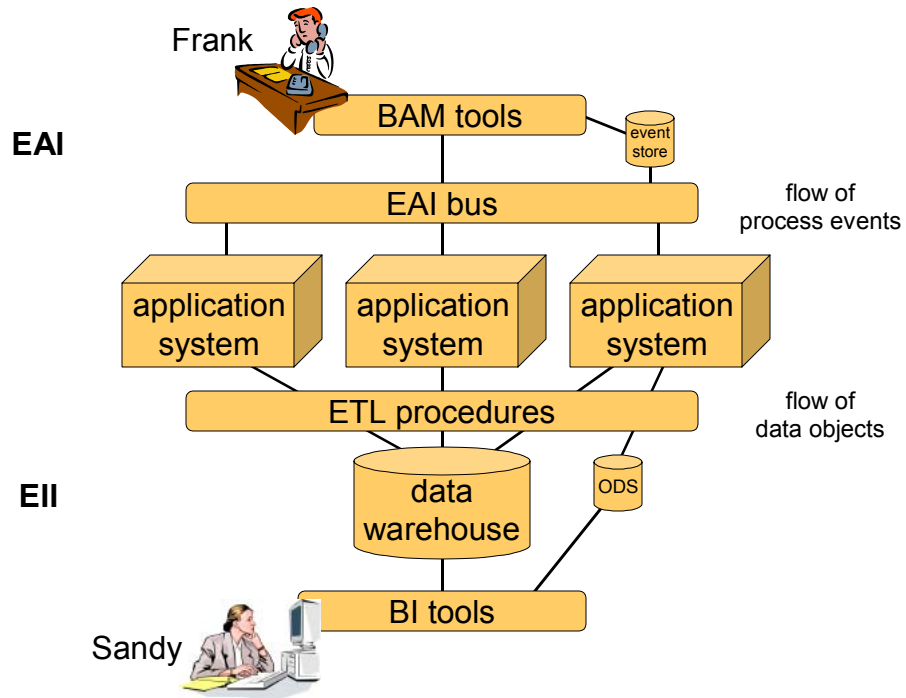


Figure 1 – Current Situation

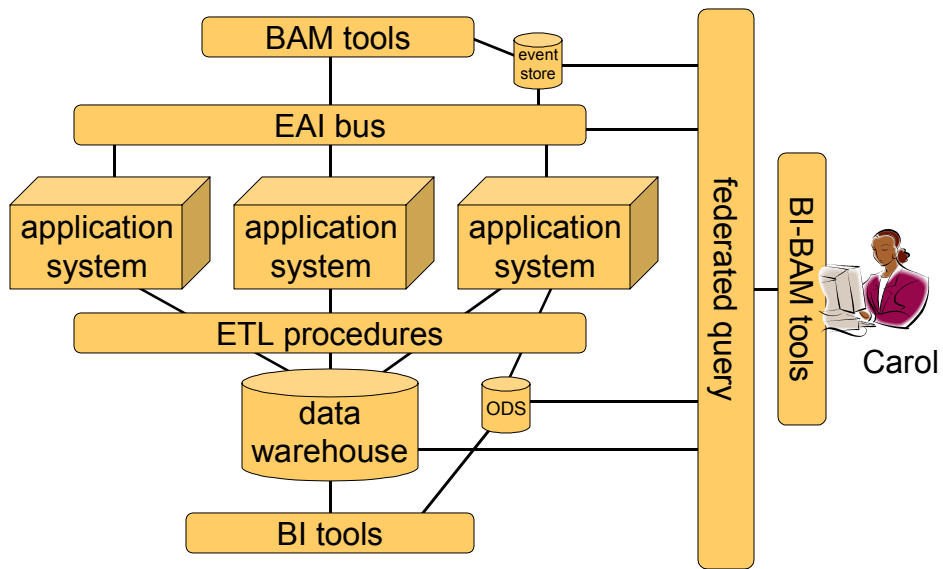


Figure 2 – Intermediate Solution