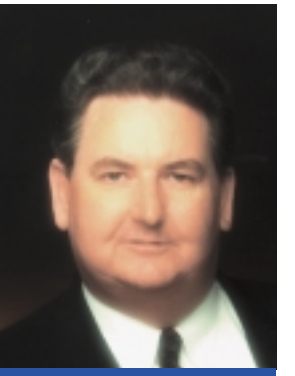
enterprise integrity



By DAVID McGOVERAN

Data Integration, Part V

Response to last month's column indicates that not everyone understands the term "data integration." Data integration is most commonly understood as data sharing, with a common database accessed by multiple applications. This month's column is a necessary aside. Before we continue our discussion of data semantics issues and strategies, we should clarify what data integration means within this series. It's one of four types of integration activities — not an integration strategy or architecture.

A first integration activity type is infrastructure integration. Infrastructure includes all the information technologies that provide operational support for applications. The plumbing metaphor is often used to describe infrastructure, but infrastructure is much more than just the interconnections between technologies. Infrastructure includes system management facilities, database management systems, storage systems, network facilities, servers, operating systems, and so on.

Though configured to support particular business operations and requirements, these facilities are business neutral. That is, the intended business operations could not be deduced simply from a list of infrastructure purchases. Experienced IT professionals know well (and may painfully remember) that achieving today's infrastructure interoperability sophistication has been a long, unfinished road of trial and error. That road is littered with superseded and failed standards, obsolete technologies, and extinct fads.

Unless the infrastructure components required for applications are integrated, those applications cannot be integrated. Infrastructure integration requires an appropriate choice of connectors, protocol converters, bridges, Application Program Interface (API) isolation layers, and more.

A second integration activity type is data integration. Because data is the carrier of all information in a business, the data produced by one or more applications must be brought into a standard, well-defined, meaningful, and consistent form if any other application is to consume it. With a data sharing strategy, the shared database removes data's transient character, but can introduce relatively long latencies between data production and consumption. It's essential to develop a common data model and interfaces to the shared database, but these tasks require relatively high upfront investments.

Recent approaches to bringing data from one application into a consumable form for another application are more incremental. For example, integrating applications point-to-point limits data integration activities to a specific data set, but usually embeds integration code, making validation and maintenance costly. It also leads to the well-known interface proliferation problem. Introducing rule-based data transformation and mapping tools can solve the former problem while hub and distributed architectures address the latter.

Unfortunately, the tendency is to forget about the need for a common data model and to downplay the importance of semantic consistency. Data integration activities, and the need to understand them, remain — no matter how little latency exists between data production and its subsequent consumption.

A third integration activity type is event integration. There are many types of events that can trigger an exchange of data among a group of applications and many ways to detect them. Writing to or reading from a database, or committing a transaction, are typical detectable events. APIs written specifically for integration may post event notifications so that adapters can capture or provide appropriate data. Event integration requires the detection of these events and the routing of event-related data among applications. This is the province of adapters, message brokers, and event brokers. Event integration need only be concerned with semantics as it pertains to connecting events in a meaningful way. Transforming the data associated with an event is a data integration task.

The fourth integration activity is process integration. Integrating business activities into a coherent business process requires a strategic, operational understanding of business objectives. Care and skill are required to capture existing business processes, and to design robust, efficient ones. Branch, merge, and loop control must be based on data that has been made meaningful and commensurate, another part of the data integration task.

Only infrastructure integration avoids direct dependence on business data integration activities. While infrastructure integration activities are primarily driven by technology changes, data integration activities are driven by business changes. It follows that all data integration errors are serious and probably costly if the data involved is important to the business. Next month, we'll return to the issues and techniques of semantic integration. Until then, consider how data integration investments offer tangible returns, and give your enterprise a chance at sustainable integrity. eAI

David McGoveran is president of Alternative Technologies, Inc. He has more than 20 years' experience with mission-critical applications and has authored numerous technical articles on application integration. e-Mail: mcgoveran@alternativetech.com; Website: www.alternativetech.com.