Getting Data to Applications—Why We Fail
Part 1: Common Fallacies

EVERYONE WANTS UNIVERSAL, TRANSPARENT ACCESS TO DATA—AND EVERYONE WANTS THE POT OF GOLD AT THE END OF THE RAINBOW. THIS TWO PART SERIES EXAMINES WHY SUCH EFFORTS GENERALLY FAIL, AND HOW WE CAN DO BETTER.

To enable better business processes and reduce duplicate data collection, our customers need to share their data more widely, both within and between their organizations. To do this, many initiatives have pursued a grand vision of "transparent access"—making all data available to all consumers (users and applications) in a way each consumer can interpret anywhere and at any time. The lack of success is often blamed on a lack of commitment and resources, but we suspect that doubling the budget would probably lead to a failure twice as costly. The deeper reasons lie in fallacious assumptions (often unconscious) on the part of the builders and their management. In the first of this two part series, we examine some of these faulty assumptions and how they can lead to failure.

Fallacy 1: The work will eventually be "finished"

Completing a task is a good thing, especially in the eyes of managers. But data integration for a large enterprise is much more like a process than a destination—continuous change is certain. Believing in an end-state where the work will be "done" tends to produce one of two problems. On one hand, it can encourage a static system architecture that meets current objectives but cannot be updated to meet new objectives. Or, it can focus on future architecture, providing no help for imperfect present day and near-term systems. Instead, since transition is a permanent condition, we need an architecture that supports incremental improvement as a permanent feature.

Fallacy 2: A single data standard is possible

Data integration requires that producers and consumers interpret the same data in compatible ways. A common approach is insisting that everyone adopt standard data models and data element definitions, either internally, or at external interfaces. It is certainly good to avoid unnecessary diversity. But no monolithic data standard will describe all systems in a large enterprise with many autonomous participants, continuous change, and 20 years of coexisting technology. Multiple standards are inevitable, so we should look for ways to both minimize the number of standards, and help system builders deal with more than one.

Fallacy 3: Data standardization is sufficient

Data standardization allows a consumer to understand data from any producer who conforms to the standard(s). However, it does not help the consumer discover which producers have relevant data, understand what portion of the desired data is actually available, or recognize and reconcile multiple reports about the same real-world entity. (Is "John Public" in one source the same as "John Q. Public" in another? If so, what if the sources disagree about his birth date?) Problems can arise in both the data definitions built into those systems, and in the data values captured. The solution must involve the people who operate the systems, not just the builders.

Fallacy 4: Mandates will accomplish what we intend

Data integration for a large enterprise necessarily involves collecting high-quality metadata (i.e., good descriptions of systems and their interrelationships) from many participants. Paper mandates will not motivate

these people to provide (and to maintain) good metadata. Mandates for metadata are typically all pain and no gain for the people who bear the costs—and while the collecting authority can check the form of the responses, they cannot check the contents. The result is predictable: metadata that satisfies the format rules but that quickly becomes obsolete and is of insufficient quality to drive operational interfaces. The solution lies in providing positive incentives (i.e., making participants' tasks easier if they cooperate).

Fallacy 5: Infrastructure funding is "nobody's problem"

Individual systems are funded because they promise immediate new functionality to end users. Infrastructure funding does not do this. However, organizations that only pay for separate systems will only get separate systems. Providing some sorts of functionality—including data integration – necessarily spans the boundaries of fielded systems and the management structures used to build them, and must be funded and managed across those boundaries if it is to succeed.

Fallacy 6: Systems should behave like wall plugs

In many discussions, we use the metaphor that we want systems to meet their data needs as if plugging into a "power grid" with a "wall plug". There is an important element of truth here; namely that each individual provider and consumer should be designed to plug into the overall system, without necessarily knowing who else will be connected. This is the only way to scale up. However, a wall plug is too simple an analogy when considering system interface requirements. A better analogy would be some of the interfaces on the back of a typical computer. They have numerous pins, and require agreements about what flows through each pin,

as seen in Figure 1. If conversion between one type of connector and another is required for a given interface, what flows through each pin must be described, and a transformation worked out for each flow. In other words, *interfaces connecting systems must be defined in terms of multiple aspects, each one of which is important.*
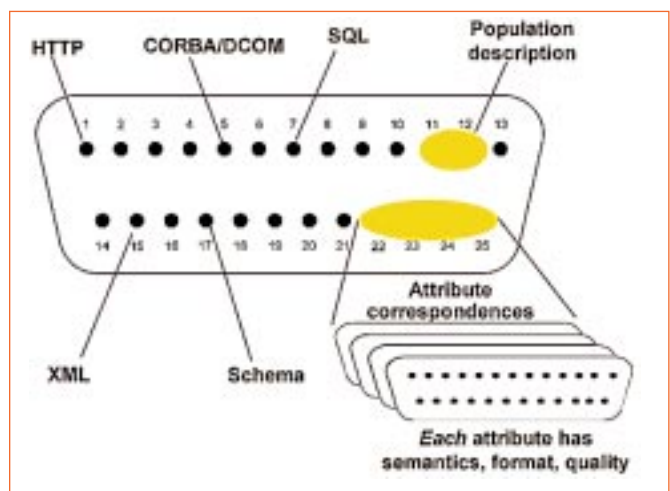

Figure 1. A better interoperability metaphor—a multipin connector. Solutions must deal with all pins

This concludes our discussion of common fallacies that inhibit data sharing across organizations. Part 2 of this series will discuss how we can improve incrementally, while minimizing risk and disruption.

For more information about improving data sharing across large, decentralized organizations, contact

Arnie Rosenthal   *Arnie@mitre.org*

Frank Manola   *fmanola@mitre.org*

Len Seligman   *seligman@mitre.org*

MITRE

www.mitre.org