BroadstreetData | Sustainable Data Intelligence

INFORMATICA® The Data Integration Company™

Untangle the IT Knot

Developing a Comprehensive Data Integration Strategy

WHITE PAPER

This edition published December 2008.

Table of Contents

Executive Summary

Many companies have allowed their IT infrastructures to grow piecemeal into a tangle of poorly integrated systems—a situation that unnecessarily complicates their ability to compete in an already complex global marketplace. At the root of the problem is the fact that companies evolve over time, as do the systems and the policies and procedures that govern them.

However, stasis is not the answer. Instead, companies need to focus on making the IT infrastructure more streamlined and efficient through more comprehensive, enterprise-wide integration.

This white paper suggests that the best approach to integration is to adopt a standardized integration platform, define an explicit integration strategy, and implement it using an Integration Competency Center (ICC). Instead of "reinventing the wheel" with every project, organizations with an ICC can optimize resources and build on past successes, leading rapidly to lower maintenance costs, more stable operations, and faster response to changing business conditions.

By the end of this white paper, you will understand:

- Why an unnecessarily complex knot of integrations is currently diverting large portions of your IT budget away from more strategic projects
- How to address the core causes of the problem
- How to rationalize and consolidate integration technologies to cut costs and leverage economies of scale
- How to choose a unified, scalable, standards-based data integration platform and apply it consistently across the IT organization

The Challenge: Balancing Complexity and Manageability

In today's uncertain macroeconomic environment, companies are employing four important strategies to operate more efficiently and reduce costs:

1. **Globalization** to diversify their operations
2. **Consolidation** to scale their operations
3. **Risk mitigation** to protect their operations
4. **Governance** to comply with a growing number of industry standards and government regulations

Each of these strategies initiates exponential growth in software, hardware, and data, which, in turn, creates a complex IT infrastructure that's difficult and expensive to manage.

Many companies make this complex environment even more complex than necessary by integrating systems one on one without adequately defining technology constraints, development standards, life-cycle processes, or maintenance documentation. This tangle of point-to-point integrations, as **Figure 1** illustrates, leads to high maintenance costs, needless delays in implementing changes, and unpredictability in operations.

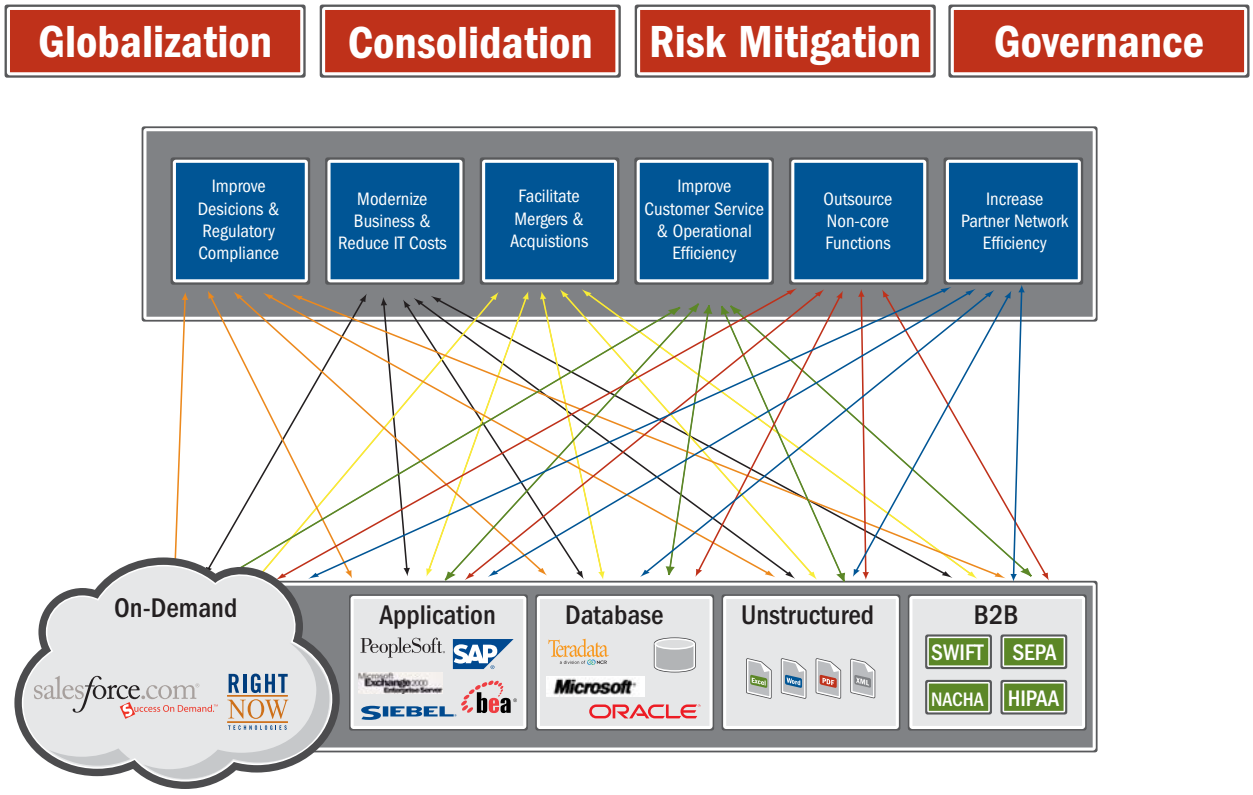

Figure 1. Unnecessary Integration Complexity

This tangle, which we call the "integration hairball," does not occur within a single system. The hundreds or thousands of components that make up a given business application are generally engineered to work tightly together according to a consistent standard developed and imposed on the design by a chief architect. Instead, we mean the interfaces among applications that were independently developed, possibly using different technologies, and which continue to be independently managed. We use the term "integration" to refer to getting these disparate applications to work smoothly together.

The problem is not the number of integrations, but the lack of standards caused by multiple independent integrations done without a master plan or sufficient attention to the ways they affect each other. This unnecessary complexity becomes particularly problematic in the typical Global 2000 company, which requires thousands or tens of thousands of integrations to support its business operations. It can divert a significant amount of your IT budget away from strategic investments for the sake of "keeping the lights on" with maintenance and ongoing operations. In fact, according to a 2007 study by Forrester Research, the average organization spends 60 percent to 70 percent of its IT budget for ongoing operations and maintenance, leaving 40 percent or less for investing in new capabilities and competitive advantage.[1] As one CIO said, "I have a billion-dollar budget and no money to spend."

This white paper will suggest how businesses can avoid unnecessary integration complexity while maximizing organizational efficiency, rationalizing the IT infrastructure, and increasing operational excellence: specifically, through the use of an Integration Competency Center (ICC). We will review the root causes of the problem, present a strategy for untangling it, and provide practical tips for executing the strategy effectively. Finally, we will address crucial questions about making the transition to a simpler, more efficient environment while still achieving these key goals:

- Supporting mission-critical projects by protecting availability and preventing data loss
- Minimizing the impact to 24x7 operational systems
- Involving new types of operational business and IT owners not traditionally included in analytical applications and other such enterprise solutions
- Increasing ROI by leveraging current resources

[1]Andrew Bartels, "U.S. IT Spending Benchmarks for 2007," Forrester Research, Inc., November 13, 2007.

Root Causes of Unnecessary Complexity

Understanding the root cause of the "hairball" allows us to attack the core problems rather than just the symptoms. These core problems include:

- Evolving technologies
- Inconsistent standards used by independent project teams
- An undefined or poorly adopted information architecture
- System-centric rather than business process-centric implementations
- Mergers and acquisitions
- Immature practices for maintaining documentation

Core Problem 1: Evolving Technologies

In the 1990s, integration tools surfaced as a new class of software, growing into an industry niche with thousands of products from almost as many vendors. As we moved into the 21st century, new products and vendors continued to emerge, but larger players began to acquire smaller ones. As we approach 2010, we are seeing a smaller number of platform suppliers with very mature offerings (this is actually part of the solution, as we will show in the next section). Nonetheless, the rapid evolution of the field has combined with traditional hand-coded integrations in a wide range of programming languages and tools to leave a legacy of multiple incompatible technologies.

Core Problem 2: Inconsistent Standards

At many companies, individual development teams have been allowed to make independent decisions that affect only their systems and meet only their specific project objectives. Over time, this has created inconsistencies and lack of standardization across teams and systems. Because every system in a given IT infrastructure interconnects in a complex web of information exchanges, the end result is suboptimal performance and sometimes even chaotic disruptions in data flow.

For example, one organization used MQ as its enterprise-wide message queue technology, but different queues had been configured by different project teams. Eventually, the rising number of production incidents persuaded the company to implement an ICC. The ICC discovered 800 servers handling 80,000 message queues—20,000 of them no longer used. Of the remaining 60,000 queues, 36,000 had inconsistent parameter settings that worked fine during normal operations but caused production outages during peak loads (e.g., inconsistent alert thresholds, maximum queue sizes too small for peak volumes, and incompatible time-out delays between dependent queues). In short, 25 percent of the infrastructure was totally unnecessary and 60 percent of the remainder was unnecessarily complex. Once the ICC applied consistent configuration rules and changed the governance controls, the MQ infrastructure went from being one of the biggest problems for the company during peak busy periods to being invisible; it simply worked.

Core Problem 3: Undefined or Poorly Adopted Information Architecture

The third source of needless integration complexity is the lack of an enterprise information architecture, or if one exists, the lack of its universal adoption across the organization. In the 1980s and early 1990s, companies commonly developed a detailed enterprise data model that could be imposed on all application components to ensure a consistent definition of information. Unfortunately, this coincided with a trend toward buying complex business systems rather than incurring the cost of building them. As companies bought applications from multiple competing software vendors, they ended up with as many data models and variations in the meaning of information as there were vendors. As a result, the enterprise data model was effectively rendered meaningless. An ICC that uses leading practices can solve this dilemma, as this paper will show.

Core Problem 4: System-Centric Rather Than Business Process-Centric Implementations

Another common error occurs from looking at each business system independently when analyzing, building, and maintaining information exchanges with other systems. This seems rational at first, assuming that each system has a detailed and thorough understanding of all information exchanges with other systems to which it directly connects. However, the whole truly is greater than the sum of its parts. As data moves among multiple systems, subtle transformations can build on each other to generate an unexpected outcome—something that theorists call "emergent behavior from complex adaptive systems." Business and accounting executives call it inconsistent data that prevents them from arriving at a single version of the truth.

Core Problem 5: Mergers and Acquisitions

When two independent organizations come together, their past independent actions will have resulted in incompatibilities. The problem is exacerbated when companies fail to eliminate the differences as soon as possible after the merger or acquisition. Within a year or two, the enthusiasm for rationalizing incompatibilities begins to fade, and the new joint organization moves on to new challenges. This piles the remaining incompatibilities atop any remaining legacy integration problems—and that keeps the "hairball" growing.

Core Problem 6: Immature Practices for Maintaining Documentation

Project teams often create documentation for new integrations using static models built into tools such as Microsoft Word, Visio, or Excel. However, this documentation is virtually unusable for day-to-day operational management, impact analysis, problem resolution, or change management. Why? First, even if all documentation is stored in a standardized place and format, it becomes hard to find the right document over time. Second, documentation standards and terminology tend to change over time, creating variations that require anyone using the documentation to relearn the integration from scratch.

Remember, the number of integrations is not the critical factor in creating unnecessary integration complexity. Suboptimal performance results from variable, undocumented, and poorly understood dependencies. To address them, you must understand what an integration system is, select an integration platform, and implement an integration strategy.

What Is an Integration System?

An integration system is a collection of components that are managed as a unit to coordinate data. It is separate from, and provides services to, other application systems within an enterprise. The services may include data migration, data consolidation, data synchronization, data quality, or process orchestration, to name a few. This contrasts with the traditional view, in which integration components are managed as part of a business application or are not managed at all (i.e., they are "orphan" integration components).

An integration system views the integration components from a holistic perspective. It defines clear boundaries around each business application and explicitly defines all the components that collectively represent the integration system and its functions, regardless of how distributed those components might be. Such a system can sustain data integration across applications after initial projects are completed.

Information exchanges between two business applications are often depicted at the conceptual (whiteboard) level as simply a line, as **Figure 2** shows.

Figure 2. Conceptual Integration View

Figure 3 illustrates how integration systems can be isolated from application systems. This is how the information exchange might look like in real life (physical level). In this simple example, the "line" in Figure 2 is an extract-transform-load (ETL) hub that directly accesses the database associated with the source system and the target system. In this example, the "line" from the conceptual view is a system in its own right, combining hardware, software, and operational processes.

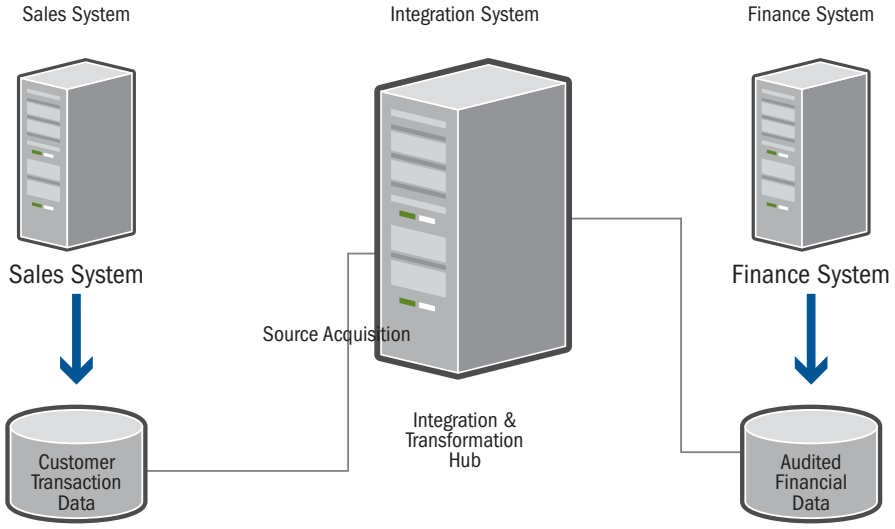

Figure 3. Simple Integration System Scenario

In a more complex scenario (**Figure 4**), the simple "line" in the conceptual Figure 3 is really a complex set of interactions between multiple components. A separate extract program is associated with the sales system. The load program for the finance system has (by agreement) been assigned as part of the integration system. Note that in this example, the queues that are part of the transport network and the security server are also part of the integration system.

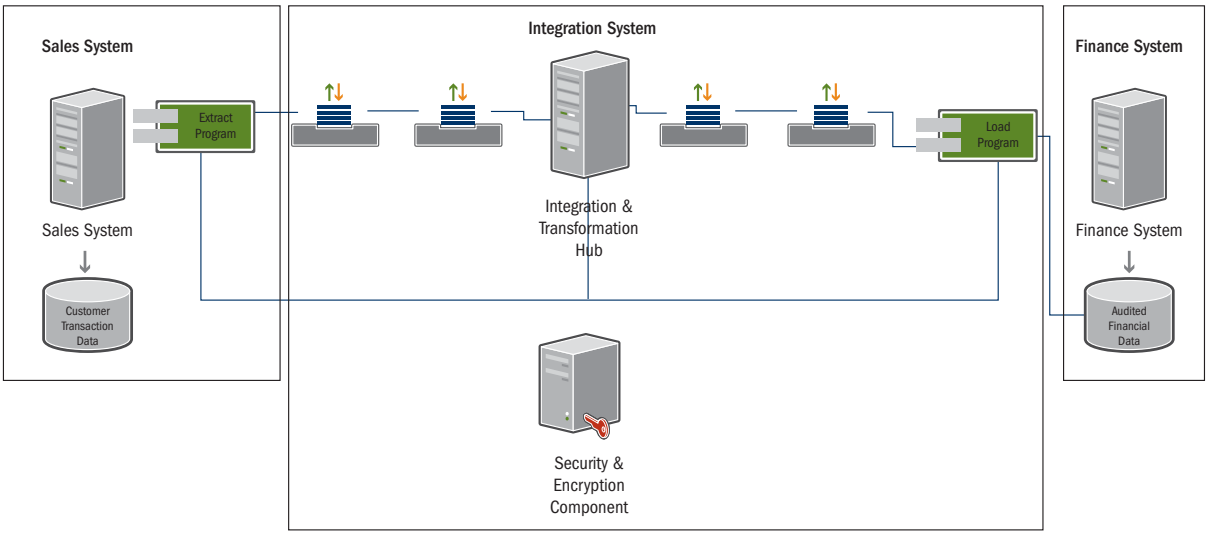

Figure 4. Complex Integration Scenario

Four questions need to be addressed for an ICC in this specific scenario:

- Is the extract program part of the sales system or part of the integration system?
- Is the load program part of the finance system or part of the integration system?
- Are all of the queues part of the integration system or just some of them?
- Is the security and encryption component part of the integration system?

Those responsible for managing, operating, and sustaining each of the components must explicitly define and agree on the answers to these questions, both to eliminate ambiguity and to assign clear accountability for all components to either the business system or to the integration system.

Selecting a Data Integration Platform

Modern integration technologies are now mature enough that every organization should establish which tools and technologies it will allow IT staffers to use to build and sustain integrations, as well as the conditions under which those tools and technologies will change. Although some amount of overlap or redundancy may be appropriate in special cases, the rule in general should be to select the fewest products needed to cover the full spectrum of integration needs. Furthermore, as integration technology continues to mature, organizations should be open to choosing new products that offer opportunities for improvement.

Figure 5 shows a framework for classifying the huge variety of available products in a way that can help minimize duplicate or unnecessary technologies.

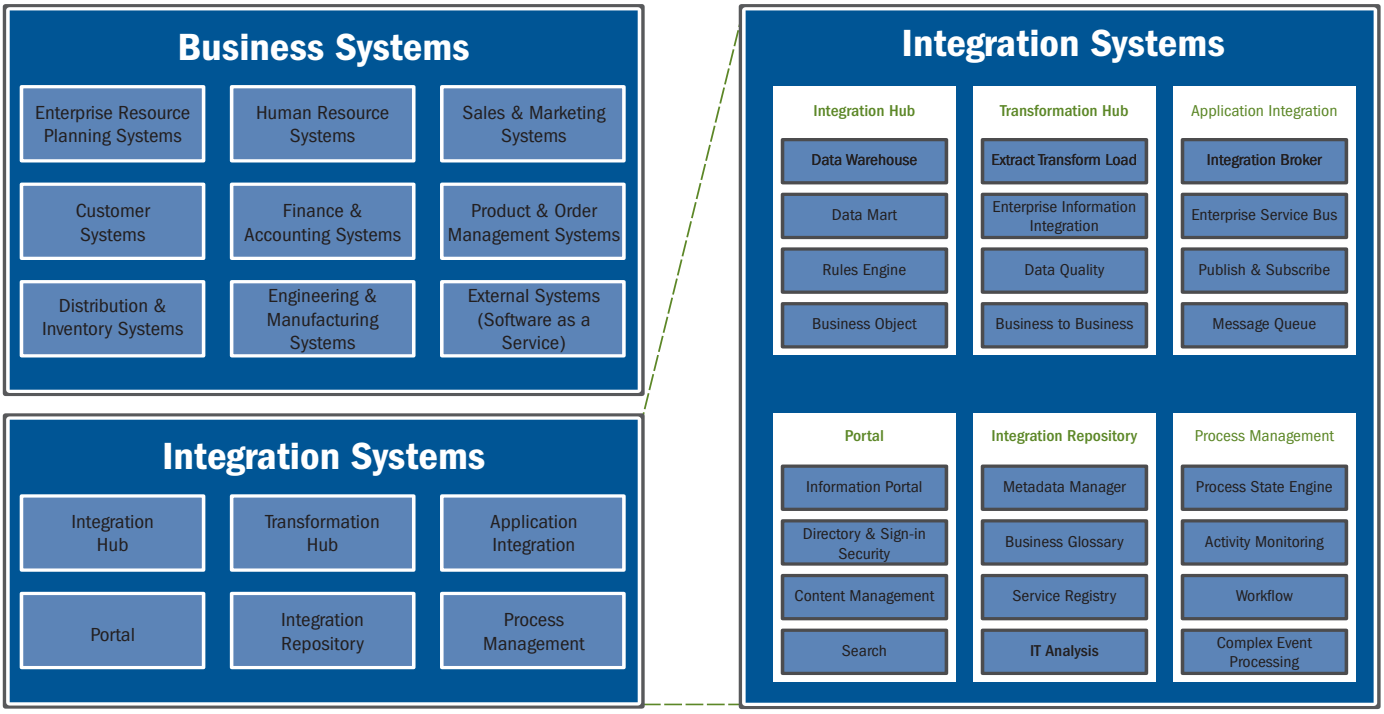

Figure 5. Taxonomy of Integration Systems

To be clear, this is a framework of integration systems—one or more products that have been configured, deployed, and run in an operational setting—and not a framework of individual technologies. The six integration systems in Figure 6 are a high-level taxonomy referred to as "systems families." They in turn decompose to a more granular list of systems.

Note that Figure 5 contrasts integration systems with business systems. Business systems support business functions such as accounting, marketing, manufacturing, and sales. Integration systems unite the individual business systems into a cohesive whole by consolidating, adapting, cleansing, and transforming data. Integration systems also control and monitor end-to-end business processes and provide an integrated experience to users or external customers. In other words, integrations between applications are viewed collectively as a "system" rather than as point-to-point appendages of the business applications.

Do not expect a one-to-one alignment between vendor products and integration systems. Many off-the-shelf products provide functionality from more than one category of integration system. For example, **Figure 6** shows which portions of the framework the Informatica® data integration platform supports and where tools from other vendors could complement the platform with business process management (BPM), service-oriented architecture (SOA), or enterprise application integration (EAI) capabilities.

Legend

BPM – Automates and coordinates the tasks that make up business processes
SOA – Provides loose coupling of application services that participate in business processes
EAI – Links applications (e.g., via API request/reply or pub/sub models) to simplify and automate transaction processing
DI – Delivers consistent, accurate, and complete data across the enterprise at any latency

Figure 6. Integration Technology vs. Integration System Map

Implementing a Data Integration Strategy

After selecting an integration platform, an organization needs to develop an integration strategy that addresses the people, process, and policy dimensions involved in its implementation. For people, the strategy should address training needs, job performance standards, organizational structure, and roles and responsibilities. For processes, it should address technology evaluation and selection, integration life cycle, alignment with project management and architecture, and documentation best practices. For policies, it should define a governance framework and integration principles and address funding issues.

Strategy is valuable only when it is executed. Therefore, above all, an integration strategy must have an organizational framework and structure that enables exceptional and sustainable execution. It needs an Integration Competency Center (ICC).

Five different organizational models may be applied depending on ICC maturity, organizational size, business operating model, and other factors, as **Figure 7** shows. The Best Practices model is most appropriate when the primary desired benefit involves maximizing the effectiveness of the IT staff. The Technology Standards model is most useful for optimally leveraging existing IT infrastructure. The Shared Services and Central Services models are best for leveraging information across the enterprise. The Self Services model is most effective in well-defined problem domains where business users distinctly benefit from the ability to build their own integrated solutions rapidly without the direct involvement of IT staff (e.g., using software-as-a-service products such as Informatica On Demand™).

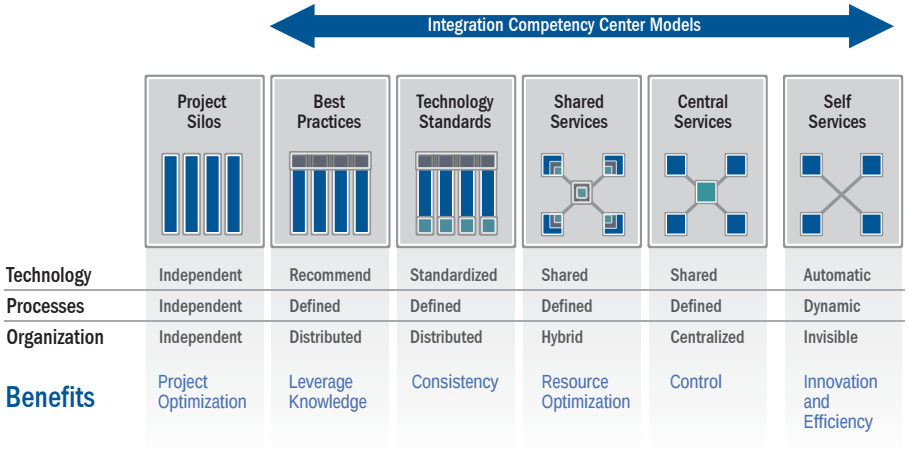

Figure 7. Integration Competency Center Models

Moving from left to right, these models leverage people, technology, and information in that order. For more detailed guidance in determining which ICC model to choose, refer to the Informatica Velocity℠ Integration Methodology Competency documentation available at https://my-prod.informatica.com.

Aside from the organizational model, six additional scope dimensions define an ICC:

1. Functional scope
2. Technology scope
3. Integration systems scope
4. Activity scope
5. ICC service scope
6. Organization scope

Figure 8 shows a framework for defining the functions of an ICC. In the broadest context, a full-functioned, mature ICC would be responsible for sustaining all internal and external integrations from initial project implementation through ongoing maintenance, support, and operations. It would have responsibility for all of these aspects:

- **Application-to-application integration** solutions, where most of the interactions between systems are managed
- **Database-to-database integration**, including a range of solutions such as batch data migration, data synchronization, data consolidation, and business intelligence
- **Business-to-business integration**, which addresses the needs of extreme loose coupling with customer and supplier systems, adaptation to industry standard protocols, and handling of semi-structured or unstructured documents
- **Business-to-person integration** (sometimes referred to as "integration as the glass"), which gives end users a view into systems through portal solutions, content management, presentation standards, and search capabilities
- **Business process orchestration integration**, including human workflow, management of long-running system-to-system flow-through processes, and coordination of processes across channels
- **Information security**, incorporating data encryption solutions, certificate management, authorization data, and single sign-on capabilities

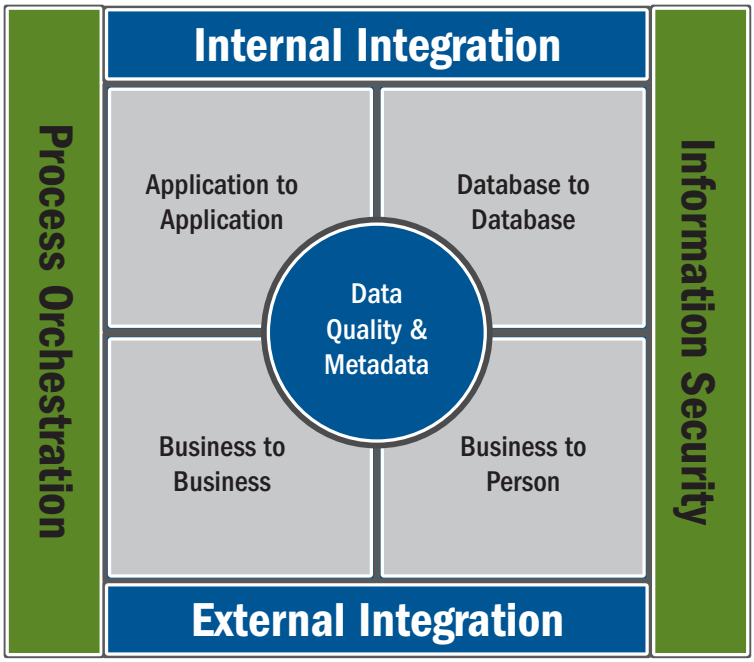

Figure 8. ICC Scope Definition Framework

As Figure 8 illustrates, data quality and metadata—arguably the most important and critical aspect of an ICC—sit at the center of the framework because they impact all dimensions of data and process integration. Data quality should be addressed at all stages of a project life cycle and in all types of integration solutions, and an ICC must have policies and practices to ensure that data is accurate and complete at point of entry.

From a metadata perspective, we are talking about managing data as an asset and documenting both "data about data" and "data about data processes." This means maintaining information about data at rest (application databases), data in motion (information exchanges between systems), and data processes such as business usage (business glossary), governance (policies and controls), and the coordination of data integration tasks.

The ability to inventory and document interfaces systematically, as part of an iterative process, is the first step to unraveling the tangle of integrations and rationalizing the integration infrastructure. After all, if you don't know what integrations exist in the organization, you can't establish improvement goals or measure progress.

Benefits of the Strategy

We are often asked, "Are the benefits of an ICC real?" and "Is this approach applicable in my organization?" The short answer to both questions is yes.

"The Economics of ICCs," an Informatica white paper, documents four specific organizational success stories. In one case, an organization that initially spent $30,000 and 30 days per interface on integration development cut those figures to $10,000 and 10 days per interface after one year, $5,000 and 5 days after two years, and, ultimately, $1,000 and a single day after three years. These results are not magic; they're the result of treating integration as a repeatable process, measuring the results, and committing to continuous improvements.

"The Economics of ICCs" also documents a company that cut its maintenance costs in half after consolidating multiple integration systems performing similar functions. The ICC led the effort to develop a business case for consolidating 30 integration systems down to 1, in the process reducing maintenance costs from $50 million to $25 million annually. Although this was a large organization with large numbers of integrations, many other enterprises have also documented similar drops in costs. Furthermore, these case studies are far from statistical oddities, as a group of Gartner analysts indicated in mid-2008:

> ICCs can save an average of 30 percent in integration application and data interface development time and costs, and 20 percent in maintenance costs, and achieve 25 percent reuse of integration components.[2]

The key message is that many, if not most, organizations have a tremendous amount of redundancy and duplication in their infrastructure and can cut costs significantly by eliminating them using an ICC with a sustainable integration strategy mission.

Note, however, that a single ICC is rarely responsible for the full scope of integrations. More commonly, organizations either establish an ICC to focus on a single area that presents a specific opportunity or a high-priority need, or they implement several ICCs with different responsibilities that do not overlap.

[2]Paolo Malinverno, Benoit J. Lheureux, Jess Thompson, and Roy W. Schulte, "Cost Cutting Through the Use of an Integration Competency Center or SOA Center of Excellence," Gartner Inc., April 4, 2008.

Conclusion

The traditional project-based approaches to data integration create the "integration hairball," but a sustainable integration strategy can untangle it, particularly if it's headed by a competency center that can take the lead in leveraging people, technology, and information across the enterprise.

With these elements as a backdrop, how then does an ICC achieve the key goals presented earlier in this paper?

- **Support mission-critical projects by protecting availability and preventing data loss.** By treating integration systems as distinct entities that are managed separately and with their own life cycles, the organization can build the initial integration for a mission-critical project more quickly and cost-effectively and then sustain the integrations over time.

  Furthermore, by viewing the collection of integrations as a single system rather than independent point-to-point interfaces, it becomes easier to implement robust, shareable infrastructures that are highly available and do not rely on single points of failure. Eliminating orphan integration points in the end reduces overall risk.

- **Minimize impact to 24x7 operational systems.** An ICC abstracts operational systems' interface specifications, captures the relevant details in a metadata repository, and implements the transformation and integration rules in a separate system. This prevents excessive interference with the core operational systems, even as business processes change or other business systems in the enterprise evolve.

- **Involve new types of operational business and IT owners not traditionally included in analytical applications and other such enterprise solutions.** The scope of the ICC must be clearly defined in terms of the services it offers, the organizational groups it supports, the stakeholders it is accountable to, and the technologies it is responsible for. Having a formal method for defining, measuring, and communicating these capabilities allows the ICC to involve a broader spectrum of the enterprise in a positive value-added fashion.

- **Leverage current resources and increase ROI to the business.** The ICC is all about reuse of assets. The assets may be people (e.g., staff in the ICC), technology, infrastructure, metadata, or the actual business instance data. The net result is the ability to accomplish objectives at lower cost, while at the same time identifying new opportunities to reuse assets that might not have been previously apparent.

Learn More

Learn more about the Informatica data integration platform. Visit us at www.informatica.com or call 800.653.3871.

About Informatica

Informatica enables organizations to gain a competitive advantage in today's global information economy by empowering them to access, integrate, and trust all their information assets. As the independent data integration leader, Informatica has a proven track record of success helping the world's leading companies leverage all their information assets to grow revenues, improve profitability, and increase customer loyalty.

BroadstreetData | Sustainable Data Intelligence

INFORMATICA®

Worldwide Headquarters, 100 Cardinal Way, Redwood City, CA 94063, USA
phone: 650.385.5000 fax: 650.385.5500 toll-free in the US: 1.800.653.3871 www.informatica.com

Informatica Offices Around The Globe: Australia · Belgium · Canada · China · France · Germany · Japan · Korea · the Netherlands • Singapore • Switzerland • United Kingdom • USA

6898 (12/15/2008)