

June 15, 2011

# Data Virtualization Reaches Critical Mass

by Brian Hopkins  
for Enterprise Architecture Professionals



June 15, 2011

## Data Virtualization Reaches Critical Mass

Technology Advancements, New Patterns, And Customer Successes Make This Enterprise Technology Both A Short- And Long-Term Solution

by **Brian Hopkins**

with Alex Cullen, Mike Gilpin, Boris Evelson, Gene Leganza, and Mackenzie Cahill

### EXECUTIVE SUMMARY

Firms are looking to solve tough data and process integration challenges as they once again begin to invest in new business capabilities. Multiple data warehouses (DWs) and business intelligence (BI) tools, rogue data marts, excessive data replication, and generally high integration costs plague enterprises; firms cannot wait for boil-the-ocean fixes. Driven by new capabilities and fueled by customer successes, data virtualization delivers on the promise of information-as-a-service (IaaS) by enabling tactical solutions that also deliver a stepping stone to enterprise data management. Introduce this technology into your integration tool kit.

### TABLE OF CONTENTS

- 2 **Firms Demand Affordable Integration Architectures With Enterprise Benefits**
  - 3 **Data Virtualization Offers A Solution, But Many Firms Have Not Adopted It**
  - 5 **New Capabilities And Usage Patterns Make Virtualization An Enterprise Option**
  - 7 **Customer Success Stories Will Fuel Further Interest**
- RECOMMENDATIONS
- 8 **Solve Integration Problems Now As You Work On The Enterprise**
- WHAT IT MEANS
- 9 **Information-As-A-Service Will Find A Broader Audience**

### NOTES & RESOURCES

Forrester interviewed several vendor and user companies as well as examined existing survey data.

#### Related Research Documents

["The Forrester Wave™: Information-As-A-Service, Q1 2010"](#)

February 10, 2010

["Information Fabric 2.0: Enterprise Information Virtualization Gets Real"](#)

April 9, 2007

["Information Fabric: Enterprise Data Virtualization"](#)

January 9, 2006

## FIRMS DEMAND AFFORDABLE INTEGRATION ARCHITECTURES WITH ENTERPRISE BENEFITS

Firms view data as an important component of business strategy — and economic recovery means they are investing more in data management solutions. According to Forrester's Forrsights Software Survey, Q4 2010, 70% of the surveyed software decision-makers are planning on expanding or maintaining their current level of investment.<sup>1</sup> More data management needs mean more data integration to deliver the knowledge executives are after.

### Data Integration Using ETL Or DBMS Consolidation Is Costly And Time-Consuming

Traditional approaches to data integration involve one of two approaches that are often not viable for short-turnaround integration needs:

- **Integration by ETL creates data quality problems and delays information delivery.** Extract, transform, and load (ETL) approaches require one or more copies of data staged along the physical integration process flow. Creating, storing, and manipulating these copies can be complex and error prone. Further, this processing is done in batches, adding a delay in delivering information updates to end reporting stores.
- **Integration by DBMS consolidation is high impact, expensive, and risky.** Consolidation of multiple database management systems (DBMSes) into a single system of record often involves extensive changes to the application environment, which affects end users. Semantic reconciliation of database schema, process, and rules integration takes time and resources as well as creates risk that the consolidated environment will not completely satisfy its constituent user groups' requirements.

### EAs Must Seek Alternative Approaches That Can Meet Tactical Demand

While ETL and DBMS consolidation remain mainstay integration tools, EAs must understand that:

- **MDM, data governance, and metadata management are long-running efforts.** These initiatives typically require high levels of attention and investment before realizing benefits. Master data management (MDM) projects typically run 12 to 24 months, and they achieve an average payback in 30 months.<sup>2</sup> While these long-term, high-benefit efforts are often necessary to achieve long-term strategic business goals, demand for better information usually cannot wait until the project's completion.
- **New integration requirements, such as cloud, are arising.** Integration gets even more complex as firms integrate software-as-a-service (SaaS)-hosted data into environments. Web-service-based integration is not the complete answer. For example, many firms plan to use Simple Object Access Protocol (SOAP) or the command line interface (CLI) to pass data between salesforce.com and their internal systems only to encounter performance issues and batch delays. A general-purpose, SQL-compliant interface is desirable in this case, but salesforce.com does not provide one.

- **The business will find a way to solve its own problems.** Absent an enterprise architecture for integration that addresses short-term demand, the business will find workaround solutions to meet silos of demand, often leading to further expense and data quality problems. Proliferation of spreadsheets, Microsoft Access databases, and copy-and-paste data movements are examples — and we are all too familiar with their impacts.

### DATA VIRTUALIZATION OFFERS A SOLUTION, BUT MANY FIRMS HAVE NOT ADOPTED IT

Data virtualization is a technology that abstracts, transforms, federates, and delivers data taken from a variety of heterogeneous information sources. It allows consuming applications or users to access data from these various sources via a request to a single access point.<sup>3</sup>

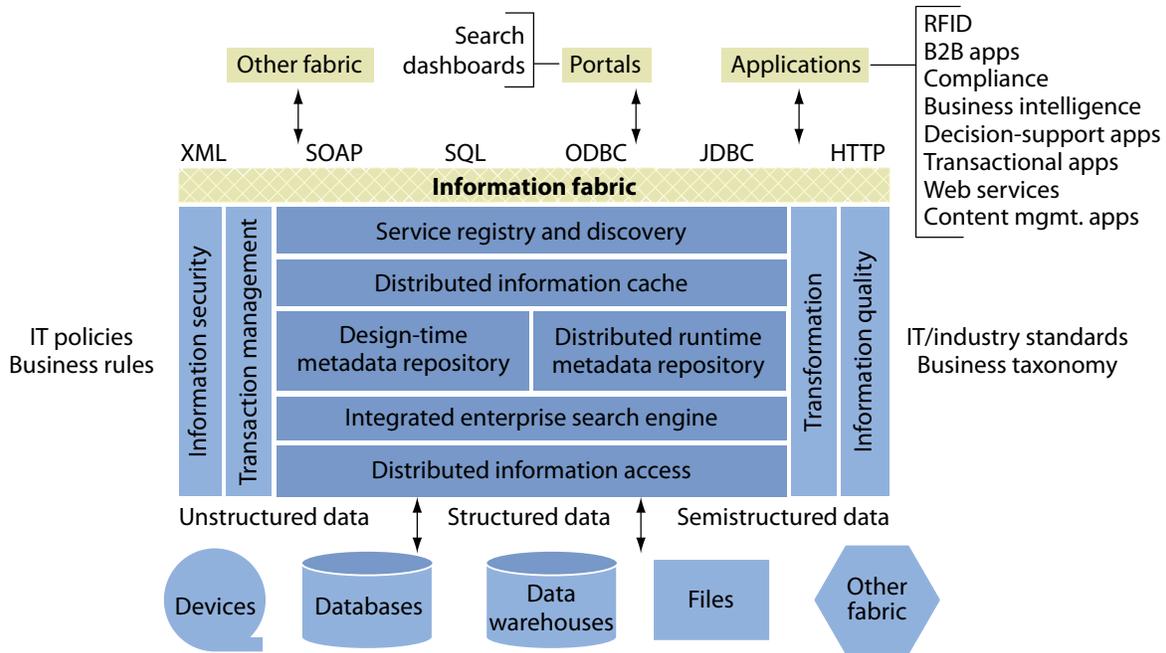
### The Information Fabric Architecture Delivers Data Virtualization For The Enterprise

Vendors such as Composite Software, Denodo Technologies, IBM, Informatica, and Red Hat all offer leading information-as-a-service solutions that implement the technology as part of an information fabric architecture.<sup>4</sup>

Forrester's information fabric reference architecture depicts the stack of technology capabilities physically required to unify various input sources and expose an output data store that is different from the sources and usually aligned with business semantics (see Figure 1). The key characteristic of this technology is that, unlike typical ETL operations, it does not actually physically move the data.

The benefits of this architecture are well documented and include: 1) flexibility and agility of integration due to the short cycle creation of virtual data stores without the need to touch underlying sources; 2) improved data quality due to a reduction in physical copies; and 3) improved usage through creation of subject-oriented, business-friendly data objects. Furthermore, standard implementation patterns exist for a wide range of project-focused requirements, including data mart consolidation, data warehouse extension, and data sharing through federation of multiple source systems.

**Figure 1** The Information Fabric Reference Architecture Implements Data Virtualization



Source: April 9, 2007, "Information Fabric 2.0: Enterprise Information Virtualization Gets Real" report

59322

Source: Forrester Research, Inc.

### More Firms Will Adopt This Technology As Pressure To Deliver Builds

Despite the technology's growing maturity and well-documented benefits, Forrester estimates that less than 20% of IT organizations have incorporated data virtualization technology into their integration tool kits and even fewer are realizing its true potential. This is largely because: 1) many early product releases haven't lived up to expectations; 2) some major vendors have underplayed data virtualization's significance or chosen not to pursue it at all; and 3) many enterprises focus on project-specific benefits, which may not justify the cost of an enterprise data virtualization platform.<sup>5</sup>

Over the next 18 to 36 months, we expect this market attitude to change as technology advancement, more third-party integration, and new usage patterns lead to increasing awareness of data virtualization's potential. Already, many early adopters are having significant success with recent releases of the market-leading products. For example, one interviewee from Qualcomm stated: "When we realized that we didn't have to physically move data around for integration, the technology started to really make sense. Now we have gone from point solutions to an enterprise deployment [of data virtualization]."

## NEW CAPABILITIES AND USAGE PATTERNS MAKE VIRTUALIZATION AN ENTERPRISE OPTION

There is a shift in client attitudes toward this technology away from seeing it as a point solution for project integration and toward treating it as an enterprise data management approach that can be implemented tactically as well as strategically. Two factors are contributing to this trend: technology advancements and the implementation of layered architectures with a canonical mapping at the center, which we refer to as an hourglass architecture.

### Technology Advancements Address Shortcomings And Deliver New Value

Inability to meet performance requirements, difficult configuration, and lingering security concerns have been the significant contributors to vendor products' failure to meet user expectations.

Technology advancements are changing this:

- **Cost-based query optimization increases the number of usage patterns.** All leading vendors spend a significant part of their research budget improving query performance, and most now feature sophisticated rule-based query engines that optimize sequencing based on expected server resource cost models. Some vendors we spoke with reported improvements in query response times ranging from 50% to 150%. As a result, though performance requirements have been an issue in the past, now performance improvements have made more options available for implementing the technology.
- **Distributed caching enables enterprise-scale operations.** Data caching is the second major contributor to performance improvement, and distributing caching operations allows cache data to reside closer to calling applications. Vendors are delivering distributed caching integrations with enterprise messaging software to deliver even better performance. For example, Composite's distributed edge caching capability allows incremental updates to cache data using IBM MQ Series to manage cache updates across several geographic locations, reducing network traffic requirements and increasing query performance.
- **Improved discovery tools make virtual data stores easier to create.** Historically, creating a virtual data store has required intimate knowledge of the source schema, which is often poorly documented. Most leading vendors have added data discovery or profiling tools to enable exploration of potential data sources; for example, Informatica allows analysts and IT users to do unique midstream profiling on federated data, in real-time, eliminating the need to pre- or post-process or stage the data for profiling, thus accelerating the process. Discovered metadata is automatically moved to a modeling and mapping workbench, where data analysts create transformations from source to target virtual stores. It's now possible to finish a simple deployment of the technology in a matter of weeks.<sup>6</sup>

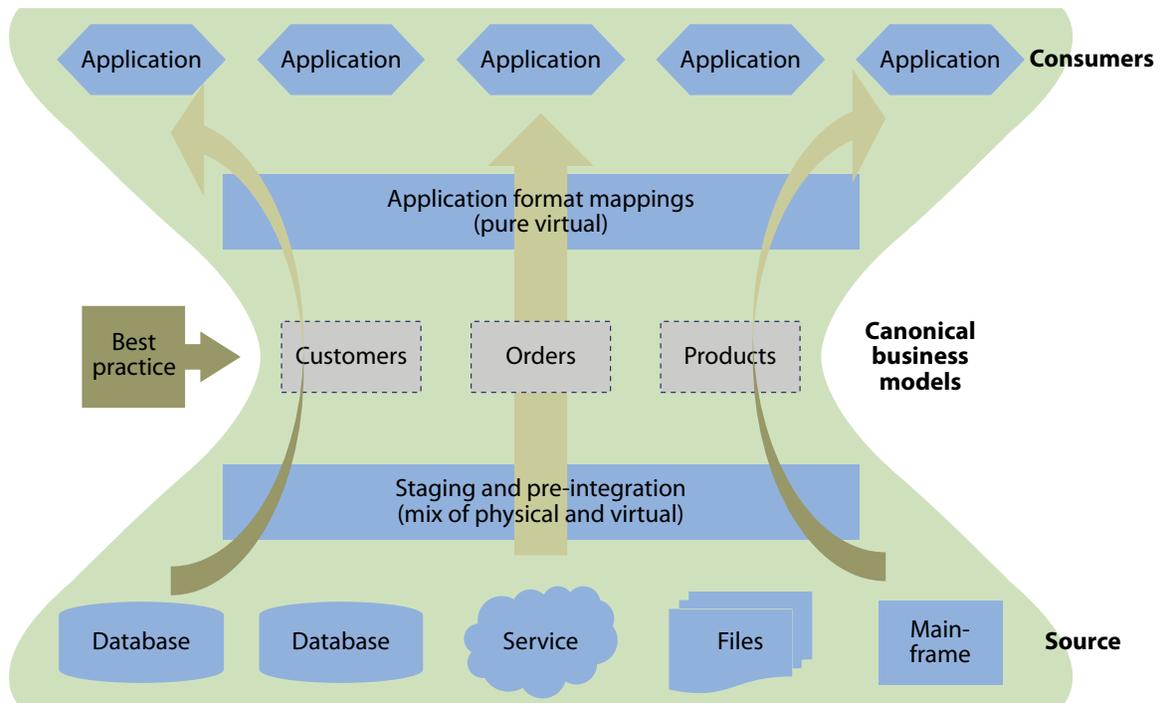
- **Data masking adds element-level protection to virtual data sources.** Virtualization technology is already capable of passing user credentials from virtual to underlying physical data stores, providing record-level security when application authorization architectures support it. An emerging capability for masking values in virtual data stores delivers additional benefits as firms expose this functionality in their data services.
- **More out-of-the-box third-party integrations create a true enterprise platform.** Aiming for competitive differentiation, vendors continue to add prebuilt integrations to their packages. As a result, data virtualization packages have become easier to integrate with the rest of an enterprise's standard technologies. For example: 1) Composite offers integrations for IBM WebSphere MQ and Tibco Software enterprise service bus (ESB) middleware that helped one company develop a global caching scheme to improve virtual data performance; 2) Informatica provides a broad set of industry standard transformations (e.g., Association for Cooperative Operations Research and Development [ACORD], Health Insurance Portability and Accountability Act [HIPAA], etc.) to jump-start data virtualization projects; and 3) integrations with SaaS and platform-as-a-service (PaaS) providers expand virtualization into cloud architectures.<sup>7</sup>
- **Integration of big data expands the potential for business insight.** More firms are recognizing the potential value locked away in massive data sets and turning to big-data processing technology, such as the Apache Hadoop framework, to sift through it all. Leading virtualization vendors recognize this need and are adding big-data adaptors to their tool kits to expose reduced sets of information gleaned from a sea of poorly structured or unstructured information.

### Clients Find That Using An Hourglass Architecture Is A Data Integration Best Practice

Leading firms have implemented a layered architecture combining both physical and virtual data stores, choosing the appropriate mix based on different areas' performance requirements. In the most-successful cases, the firms create an hourglass-shaped architecture that funnels mappings of disparate source data through canonical business information models.<sup>8</sup> As a result, one large drug manufacturer is able to successfully operate with both IBM Cognos and SAP BusinessObjects BI tools for different customer groups, using data that has been reconciled against common metadata using virtualization (see Figure 2).

In addition to the use of canonical models in the middle, we have identified two other noteworthy characteristics of this architecture: 1) virtual/physical modality choices tend to be more physical in the staging layers close to the actual data sources and more virtual as data moves closer to the users, and 2) a final virtual mapping layer ensures that the solution provides data to consumers in just the required format.

**Figure 2** An Hourglass Architecture Integrates Physical And Virtual Layers



59322

Source: Forrester Research, Inc.

## CUSTOMER SUCCESS STORIES WILL FUEL FURTHER INTEREST

As users gain experience with project-focused solutions and as the technology improves, leading firms are beginning to leverage the information fabric architecture to deliver enterprise benefits. Additionally, initiatives to integrate with the cloud are common, and big data integration is beginning. Some success stories include:

- Pfizer enables multiplatform BI.** The typical pipeline for drug production is a long, expensive process involving many business units; it averages 12 years in length and more than \$1 billion in costs. Every step of the process involves data, and different business units have their own BI technologies to perform the required analytics. For Pfizer, moving to a single BI platform was simply not cost effective; however, operating with several platforms presented integration challenges that affected quality, speed, and cost. The drug company implemented virtualization technology from Composite Software early in the integration process to: 1) prototype new data sources prior to adding them to its enterprise data warehouse, and 2) achieve better integration of data with existing BI tools. As a result, it significantly reduced the integration cycle time as well as its cost.

- **HealthNow has unified multiple data sources and done away with special-purpose data marts.** HealthNow, one of New York's leading healthcare providers, reached a crossroads in its quest to lower operational costs, reduce complexity, and speed up data delivery. It had healthcare data spread across 16 enterprise databases and 30,000 Microsoft Access databases; to meet new data needs, it built special-purpose operational data stores or marts, which took months. To solve its problems, the firm chose to adopt an enterprise data services solution using the Informatica 9 platform. As a result, HealthNow is able to rapidly process more than 10 million rows from its enterprise databases and deliver subsecond responses to queries. It also reduced the time to create virtual data stores in response to changing business requirements from months to days and realized new agility: "If a data source changes, we just change which data source the data service points to without affecting the consuming application."
- **Qualcomm started with a point solution and then moved to an enterprise deployment.** Qualcomm, a large digital wireless telecommunications products and services company, initially procured virtualization technology to access customer data in salesforce.com and combine it with ticket information from internal systems. It implemented a system leveraging vendor-supplied out-of-the-box integration to create a relational virtual database within its firewall. This provided better performance and was much simpler to implement than a straight web service approach; the company accomplished the project in a matter of weeks. Based on this success, Qualcomm has now deployed an enterprise information fabric architecture using the hourglass strategy.

## RECOMMENDATIONS

### SOLVE INTEGRATION PROBLEMS NOW AS YOU WORK ON THE ENTERPRISE

Forrester believes that integration problems need immediate solutions, and enterprise architects should not relegate their leadership to focus only on long-term, strategic efforts. By driving an information fabric architecture into integration environments, organizations can support MDM, data warehouse consolidation, and multiplatform BI as they move toward enterprise information management. To take action now, enterprise architects should:

- **Use a catalog of unmet data integration demand to introduce the technology.** Talk to your data integration technology specialist, and create an inventory of small projects that are on hold because of priority and time. Evaluate the requirements against the information fabric reference architecture and ask, "Do we need to move data via ETL here?" Build a business case, or leverage innovation funds to demonstrate technology value for good candidates.
- **Redraw the data integration conceptual architecture using an hourglass shape.** The true benefit of data virtualization is an enterprise information fabric deployment that leverages a canonical model for primary business data entities. Incorporate this concept into future state information architecture blueprints, and use a virtualization layer for the final mappings between the canonical data and application-specific forms.

- **Develop integration patterns that expose information-as-a-service.** As you successfully deploy virtual data sources, develop service-oriented architecture (SOA) integration patterns that solution architects can easily leverage in projects. This will drive adoption of the technology in the context of your larger IaaS strategy.

## WHAT IT MEANS

### INFORMATION-AS-A-SERVICE WILL FIND A BROADER AUDIENCE

In our report “The Top 15 Technology Trends EA Should Watch: 2011 To 2013,” we predict that information-as-a-service will go beyond data integration to support requirements for single version of the truth, real-time BI, and high-performance transactional applications.<sup>9</sup> As the technology becomes an enterprise norm:

- **Some firms will reduce their data warehouse footprint — or at least slow growth.** Some firms are already asking, “Do we need build out the DW, if we can grab some data from it, mash it with some other data, and rip out a virtual database in a third of the time and for a third of the cost?” The gist is this: Why continually expand and bloat out an enterprise data warehouse (EDW) that is already aging and fat until you really know that a new information set is of importance to the enterprise? In many cases, firms will choose to not move data into the warehouse where it is not warranted.
- **Virtual MDM may heat up.** The single-version-of-the-truth benefit of virtual data stores naturally lends itself to a single version of master data. Some master data sets are small enough and static enough to be a natural fit for basic virtualization, and advancements in the technology, including write-back, will only make this better. Information indicating actual progress toward virtualizing MDM is scarce, and the term is not currently buzzing. As enterprise implementations become more common, it may.
- **Integration costs will go down as application delivery speeds up.** Often, just the mention of the word integration in the context of a project’s requirements sends up the cost alarm bells, leading some solution architects to work around problems with anything but ETL. By adding data virtualization into your architecture tool kit, we believe that you can substantially drive down integration costs and speed up project delivery. As an example, clients have used this technology to transform source data into forms that more easily plug into existing applications environments, eliminating the need to recode and retest.

## ENDNOTES

<sup>1</sup> Forrester’s Forrsights Software Survey, Q4 2010, was fielded to 2,403 IT executives and technology decision-makers located in Canada, France, Germany, the UK, and the US from small and medium-size (SMB) and enterprise companies with two or more employees.

- <sup>2</sup> According to Forrester's November 2010 Global Master Data Management Online Survey of 188 DQ-savvy IT professionals, 70% of survey respondents' MDM initiatives took more than 12 months to roll out, with a full 31% taking more than two years. Thirty-eight percent of respondents' MDM deployments occurred over four or more phases. MDM initiatives take an average of 30 months to reach payback. See the December 6, 2010, "[Master Data Management Adoption Trends](#)" report, and see the October 29, 2008, "[The ROI Of Master Data Management](#)" report.
- <sup>3</sup> Forrester differentiates data virtualization from the notion of data fabric and information-as-a-service, even though these terms are often used synonymously. Data virtualization refers to the core technology stack, and data fabric is used to indicate an enterprise deployment of data virtualization and supporting technologies, whereas IaaS refers to a broadly scoped data services deployment that includes data fabric and other SOA technologies such as ESBs.
- <sup>4</sup> Forrester's 2010 evaluation of the IaaS market identified IBM, Informatica, Composite Software, And Denodo Technologies as IaaS leaders, with Microsoft and Red Hat close behind. See the February 10, 2010, "[The Forrester Wave™: Information-As-A-Service, Q1 2010](#)" report. In January 2006, Forrester identified the information fabric approach as being an enterprise data virtualization approach. See the January 9, 2006, "[Information Fabric: Enterprise Data Virtualization](#)" report. A follow-up report further defines the reference architecture stack. See the April 9, 2007, "[Information Fabric 2.0: Enterprise Information Virtualization Gets Real](#)" report.
- <sup>5</sup> Since its acquisition of BEA, Oracle has decided not to participate in the virtualization market and is not investing significantly in the data virtualization technology gained through its acquisition. IBM continues to be a leader in the broadly defined information-as-a-service technology stack market. However, it does not emphasize the virtualization aspect of its information server technology suite. See the February 10, 2010, "[The Forrester Wave™: Information-As-A-Service, Q1 2010](#)" report. Forrester estimates the cost of a data virtualization deployment to be between \$250,000 and \$500,000 for initial deployment of software and services, the high-end estimate being for enterprise deployments of an information fabric. Follow-up projects then implement the foundation across various systems and platforms.
- <sup>6</sup> We validated this estimate using information gathered from leading vendors such as Informatica and Composite Software as well as conversations with clients.
- <sup>7</sup> Integration with cloud vendors is a particularly important pattern. Typically, these data sources are accessed via RESTful interfaces. Virtualization helps make the data integrate more naturally into existing DW and BI environments and helps teams move away from traditional, suboptimal approaches. Lastly, not all PaaS vendors have shunned SQL access — for example, Microsoft Azure.
- <sup>8</sup> Forrester defines a canonical model as "a model of the semantics and structure of information that adheres to a set of rules agreed upon within a defined context for communicating among a set of applications or parties." Forrester has identified the practice of canonical modeling as an IaaS and SOA best practice. See the November 15, 2007, "[Canonical Information Modeling Is Key To Many Information-As-A-Service And SOA Strategies](#)" report. Canonical models are at the center of the hourglass architecture detailed in this report.

<sup>9</sup> Forrester began summarizing technology trends in 2009 to help enterprise architects create their organizations' technology watch lists. For 2010's list of top trends, we used the same criteria — impact, newness, and complexity — but modified the categories, merged related topics, added five new trends, and updated all the entries with the new year's perspective. 2010's categories? "Empowered" technologies, process-centric data and intelligence, agile and fit-to-purpose applications, and smart technology management. Also new in 2010 are the results from a survey we ran as input to the report that asked respondents to rate more than 40 technologies for impact to their organization in the next three years. See the October 14, 2010, "[The Top 15 Technology Trends EA Should Watch: 2011 To 2013](#)" report.

# FORRESTER®

Making Leaders Successful Every Day

## Headquarters

Forrester Research, Inc.  
400 Technology Square  
Cambridge, MA 02139 USA  
Tel: +1 617.613.6000  
Fax: +1 617.613.5000  
Email: [forrester@forrester.com](mailto:forrester@forrester.com)  
Nasdaq symbol: FORR  
[www.forrester.com](http://www.forrester.com)

## Research and Sales Offices

Forrester has research centers and sales offices in more than 27 cities internationally, including Amsterdam; Cambridge, Mass.; Dallas; Dubai; Foster City, Calif.; Frankfurt; London; Madrid; Sydney; Tel Aviv; and Toronto.

*For a complete list of worldwide locations visit [www.forrester.com/about](http://www.forrester.com/about).*

For information on hard-copy or electronic reprints, please contact Client Support at +1 866.367.7378, +1 617.613.5730, or [clientsupport@forrester.com](mailto:clientsupport@forrester.com).

We offer quantity discounts and special pricing for academic and nonprofit institutions.

Forrester Research, Inc. (Nasdaq: FORR) is an independent research company that provides pragmatic and forward-thinking advice to global leaders in business and technology. Forrester works with professionals in 19 key roles at major companies providing proprietary research, customer insight, consulting, events, and peer-to-peer executive programs. For more than 27 years, Forrester has been making IT, marketing, and technology industry leaders successful every day. For more information, visit [www.forrester.com](http://www.forrester.com).