IBM's Federated Metadata Management Strategy

Michael J. Blechar

Metadata management is becoming a pervasive problem as organizations implement business process management (BPM) initiatives and service-oriented architectures (SOAs). The use of "community based" repositories is a best practice, as well as the basis of IBM's metadata management strategy, but their meaningful federation will be a challenging vision to realize.

Key Findings

- IBM's federated metadata management strategy is visionary.

- Its support for four major communities of users with separate repositories is pragmatic.

- IBM is releasing a new repository for managing other types of metadata assets, including those related to architecture and applications.

- IBM must demonstrate that it can meet customers' use-case scenario needs to federate metadata across its four repositories.

Recommendations

- IBM-centric IT organizations should strongly consider using all IBM repositories on a community basis, and federated together on a more-selected basis.

- IT organizations with heterogeneous technologies should strongly consider using the IBM repositories that support the communities (which are, in turn, supported by IBM tools), while possibly using other vendors' repositories for communities that use non-IBM technology.

- Many will find it valuable to augment the IBM solutions with those from leading metadata repository vendors.

TABLE OF CONTENTS

LIST OF FIGURES

Gartner

ANALYSIS

As organizations implement SOAs and enable BPM initiatives, we're seeing a proliferation of new types of metadata, as well as an increased concern and focus on metadata management across the enterprise. There are other drivers for improved metadata management and understanding, including mergers and acquisitions, application package selection, implementation and governance, and compliancy legislation, such as the U.S. Sarbanes-Oxley (SOX) Act and Japan's Financial Instruments and Exchange Law (J-SOX), to name a couple. This complex problem affects a broad range of technologies and organizational roles, as well as information architectures.
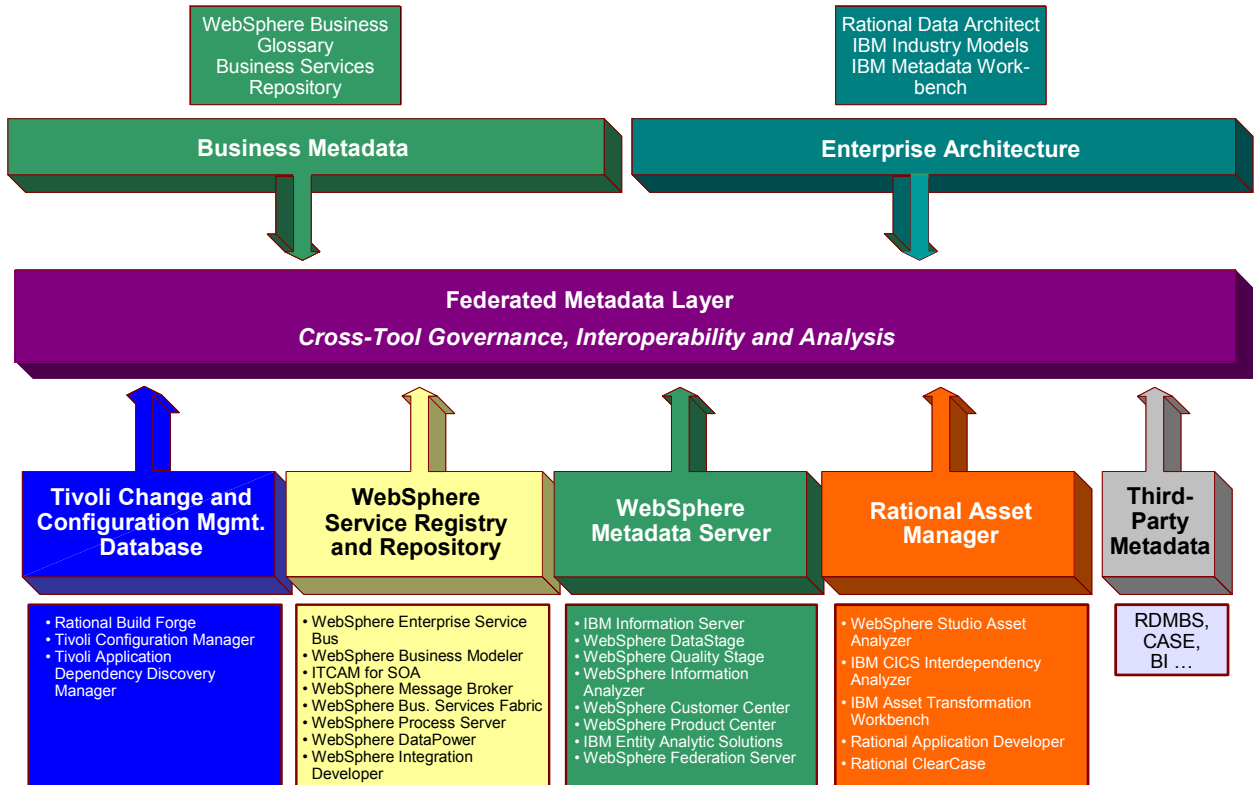
Not surprisingly, customers of the major application package and independent software vendors, such as BEA Systems, IBM, Oracle and SAP, are asking for help in addressing this problem, and are each taking different approaches to providing metadata management solutions. BEA, for example, is "OEMing" the HP Systinet tool under the name AquaLogic Service Registry. In addition, BEA acquired Flashline and its metadata tool, now called AquaLogic Enterprise Repository, and is continuing to integrate its tools and those of its partners with these products. However, in terms of the broadest scope of federated metadata repositories and registries, IBM is assuming the role of industry thought leadership with its strategy, which includes building new products focused on this approach. Its short-term, *pragmatic* focus has been to extend metadata management beyond just IBM "tool specific" capabilities and toward extended support for three "communities of users," sharing metadata about systems, data and SOAs. To this degree, IBM has already achieved the initial objectives of this part of the strategy.

In addition, IBM is adding a fourth and more-customizable way to manage other types of metadata assets, including those related to architecture and applications. We need to see how well this new community of users will be supported. It will be possible to federate metadata across the four communities of users and their repositories; however, due to the variety of metadata and tool types that must be integrated, and the disparate use-case scenarios that are possible, this is the most-ambitious aspect of the strategy, and it will need to evolve based on customer feedback.

In "The Evolving Metadata Repository Market," we examined the history of approaches to metadata management and recommended that the most-pragmatic approach was to use multiple, federated repositories and registries to handle the problem (with an optional "metadata warehousing" repository for reporting on select, abstracted levels of metadata, in case the federated approach didn't meet all user requirements). We suggested that each type of repository or registry should focus on one specific community of users and have the capability to conduct federated queries across them. For example, the community of users focused on governing runtime metadata related to workflow and services would use a different repository than a community of users involved with enterprise architecture or application development. IBM's strategy is completely aligned with this recommendation.

Gartner

1.0 IBM's Three Current Repositories

Figure 1. IBM's Metadata Vision

WebSphere Business Glossary Business Services Repository		Rational Data Architect IBM Industry Models IBM Metadata Work-bench
Business Metadata		**Enterprise Architecture**

Federated Metadata Layer
Cross-Tool Governance, Interoperability and Analysis

Tivoli Change and Configuration Mgmt. Database	WebSphere Service Registry and Repository	WebSphere Metadata Server	Rational Asset Manager	Third-Party Metadata
• Rational Build Forge • Tivoli Configuration Manager • Tivoli Application Dependency Discovery Manager	• WebSphere Enterprise Service Bus • WebSphere Business Modeler • ITCAM for SOA • WebSphere Message Broker • WebSphere Bus. Services Fabric • WebSphere Process Server • WebSphere DataPower • WebSphere Integration Developer	• IBM Information Server • WebSphere DataStage • WebSphere Quality Stage • WebSphere Information Analyzer • WebSphere Customer Center • WebSphere Product Center • IBM Entity Analytic Solutions • WebSphere Federation Server	• WebSphere Studio Asset Analyzer • IBM CICS Interdependency Analyzer • IBM Asset Transformation Workbench • Rational Application Developer • Rational ClearCase	RDMBS, CASE, BI …

Source: IBM

As indicated in Figure 1, IBM has consolidated its communities of users into four categories and tools. The first includes users who are interested in metadata regarding systems, such as those using IBM's Tivoli products. The "community repository" for these individuals is Tivoli's Change and Configuration Management Database (CCMDB).

The second category includes users who are interested in SOAs, such as those using IBM's WebSphere as an enterprise server bus (ESB), or as a workflow and orchestration server for BPM. The community repository is a product built by IBM — WebSphere Service Registry and Repository (WSRR).

The third community of users is concerned with the metadata related to "data" — that is, data architecture, data warehousing and enterprise information integration. IBM has built a product called WebSphere Metadata Server (WSMS), which includes the IBM Information Server and supports the Ascential Software tool suite. The Unicorn Solutions metadata repository, which IBM acquired in 2006, is also integrated into WSMS.

The feedback we've received from IBM customers regarding these three repositories/registries has been positive. Prior to this, there were, in many cases, individual tool-specific metadata management solutions with only a limited level of reuse via bridging mechanisms. With the new community-based repository, however, there's greater standardization and shareability of metadata across the tools used in each community. This isn't to say that IBM doesn't have more work to do in providing further standardization and integration of metadata within (and across) the

Gartner

communities; rather, most users will find that this is a positive step forward from where they were before these repositories were released.

2.0 IBM's New Fourth Repository

The fourth community of users is broader. It includes all other types of metadata — especially those related to architecture and applications (including application development and application life cycle management). The current WebSphere and Rational product suites come with some degree of integration because most are compliant with the Eclipse open-source standard. Thus, they follow a common user interface standard and can use a common set of software to conduct version control and configuration management, such as ClearCase or ClearQuest. However, to manage the types of metadata assets that the fourth community required meant that IBM had to build another product. IBM expects this product, the IBM Rational Asset Manager (RAM), to be generally available electronically by 29 June 2007.

RAM adheres to the Object Management Group's Reusable Asset Specification (RAS) standard. The initial release of RAM will focus more on IBM's design-time and development tools, but because RAS can support any type of metadata artifact, relationship or attributes, the RAM repository can be used for other needs and communities in the future.

IBM WebSphere and Rational products have stored metadata in a variety of tools and repositories for decades. Since 2000, IBM has focused on improving the level of metadata integration beyond what was previously possible. The introduction of RAM is the first step toward more-robust integrations across the IBM suite of products; however, the sheer variety of current and legacy tools in the suites means that this is an evolutionary process. It's likely to take time to roll out support for RAM on a broad technology basis, while support for less-strategic legacy technologies isn't likely to be a high priority. As a general philosophy, IBM's organizational view of the focus of its tools will probably mean that the evolution will follow this order: first, enabling individuals in communities; then the community members as a team; and then support for cross-communities and repositories.

Therefore, IBM must demonstrate that this new repository can initially meet the anticipated needs of users in the fourth community, including the requirement to support organizationally and geographically complex or dispersed groups of teams. Unlike the more-focused communities in the first three repositories, RAM is more customizable and must be able to coordinate use with a broader base of "other" communities of users. For example, it will be possible to use RAM for purposes that weren't initially the main focus of the first release, such as managing outsourcing engagements or extended compliance and governance needs. As a result, IBM customers will want to use the new tool to gain maximum advantage, but they must also be careful not to "outrun" its planned evolution.
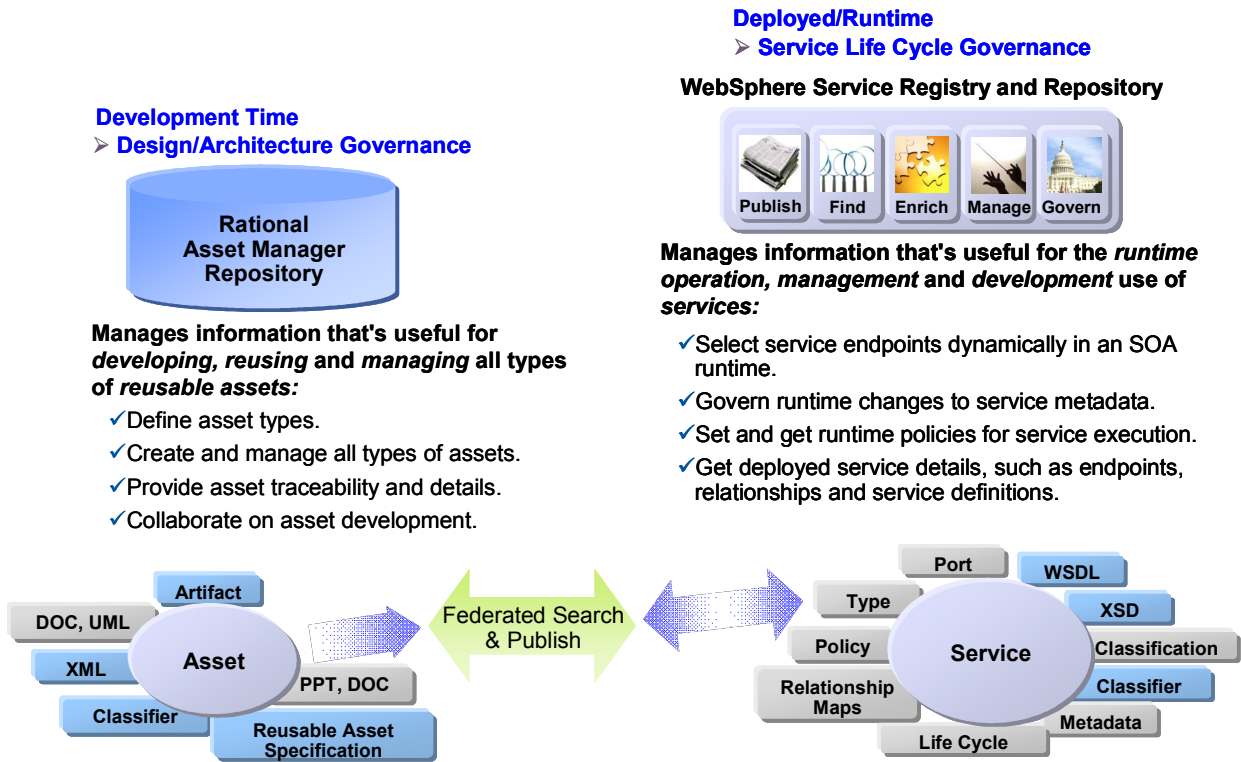
3.0 Federating the RAM and WSRR Repositories

As depicted in the Federated Metadata Layer in Figure 1, the IBM vision calls for the integration and federation of metadata residing in its four repositories and other IBM tools (such as those supporting the business and enterprise architectures), as well as technologies from third parties. However, this is an evolving strategy with regard to which metadata will reside in what repositories and how integration and federation will be conducted — a challenging endeavor.

Like most large vendors, IBM is hearing from customers that they need support for a concept we call a "business service repository" (BSR) — the governance and management of architectural and design-time metadata about business services, through the deployment of runtime software and data services. To meet a major portion of this requirement, IBM must federate two of its repositories — RAM and WSRR — as depicted in Figure 2. Arguably, IBM will also have to

Gartner

federate data and system information in the other two repositories, but a major step forward in supporting the BSR involves federating RAM and WSRR.
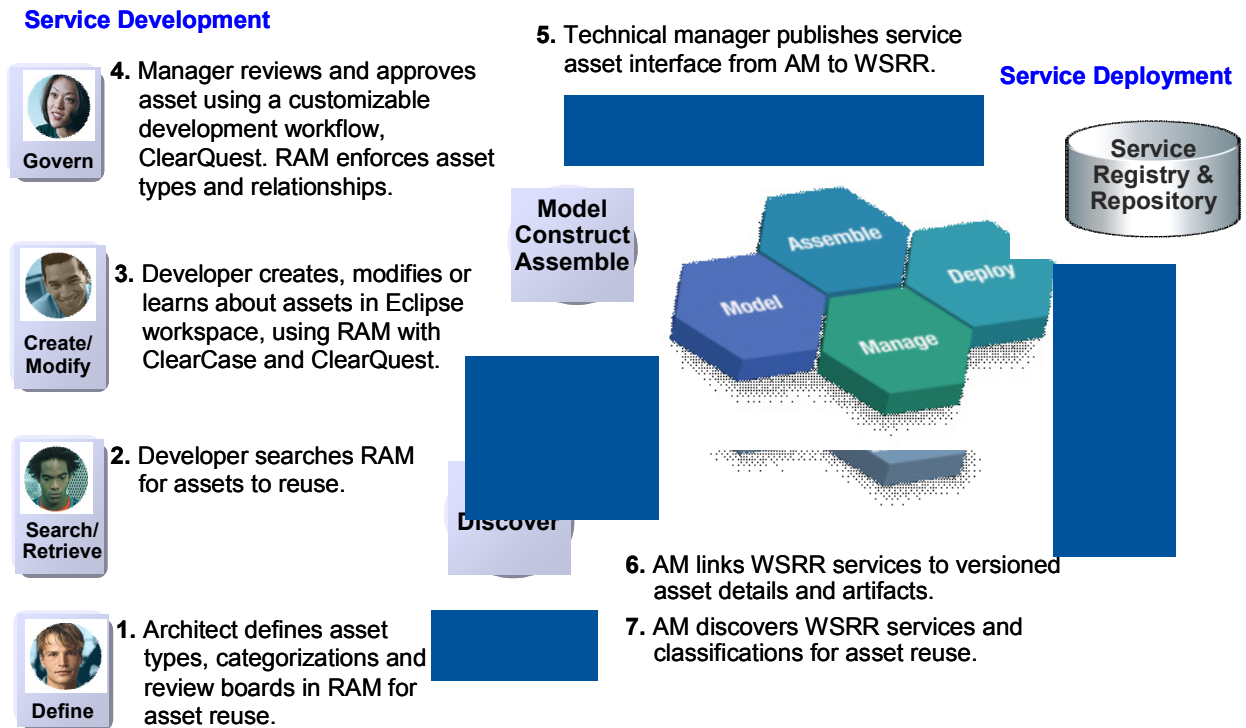
Figure 2. Federation of the RAM and WSRR Repositories

Deployed/Runtime
➢ **Service Life Cycle Governance**

WebSphere Service Registry and Repository

Publish Find Enrich Manage Govern

Development Time
➢ **Design/Architecture Governance**

Rational Asset Manager Repository

Manages information that's useful for the *runtime operation, management* and *development* use of *services:*

✓Select service endpoints dynamically in an SOA runtime.
✓Govern runtime changes to service metadata.
✓Set and get runtime policies for service execution.
✓Get deployed service details, such as endpoints, relationships and service definitions.

Manages information that's useful for *developing, reusing* and *managing* all types of *reusable assets:*

✓Define asset types.
✓Create and manage all types of assets.
✓Provide asset traceability and details.
✓Collaborate on asset development.

Artifact
DOC, UML
XML   **Asset**
Classifier
PPT, DOC
Reusable Asset Specification

Federated Search & Publish

Port   WSDL
Type   XSD
Policy   **Service**   Classification
Relationship Maps   Classifier
Life Cycle   Metadata

Source: IBM

RAM and WSRR will be integrated and federated to support a broad range of metadata types, with WSRR focused on services and RAM focused on all other types of pertinent assets. However, this "federation" isn't "magical." It requires making decisions as to where and how metadata (which must be shared or reported on) will be stored in the two repositories. It also requires the coordination of metadata ownership and use by those in different roles across the organization (see Figure 3).

Gartner

Figure 3. RAM Define, Create, Modify and Review Asset Interactions

Service Development

4. Manager reviews and approves asset using a customizable development workflow, ClearQuest. RAM enforces asset types and relationships.

Govern

3. Developer creates, modifies or learns about assets in Eclipse workspace, using RAM with ClearCase and ClearQuest.

Create/ Modify

2. Developer searches RAM for assets to reuse.

Search/ Retrieve

1. Architect defines asset types, categorizations and review boards in RAM for asset reuse.

Define

5. Technical manager publishes service asset interface from AM to WSRR.

Service Deployment

Service Registry & Repository

Model Construct Assemble

Assemble

Deploy

Model

Manage

Discover

6. AM links WSRR services to versioned asset details and artifacts.

7. AM discovers WSRR services and classifications for asset reuse.

Source: IBM

Figure 3 illustrates a sample workflow for using and federating metadata (managed in RAM) with WSRR. It also shows some of the challenges that enterprises will face when they want to go beyond their repository-specific communities of users to access more-federated types of metadata. For example, in Step 7, WSRR services are initially loaded into RAM. These might undergo changes as part of development, and if the WSRR community wants to know about these new, work-in-process versions, then they'll need to be published in the next iteration of Step 5 back to WSRR.

Decisions must be made concerning the metadata as part of the federation strategy. Deciding which metadata assets related to the production version of the service will continue to be valid, as opposed to replacing or splitting them into multiple new versions, is an analytical step that must be taken to keep WSRR accurate. The more versions of WSRR service assets in RAM that are changing, the greater the analytical effort will be in keeping WSRR completely up-to-date.

Note: As IBM evolves its metadata management rollout, it will be recommending how to use RAM with CCMDB and WebSphere Metadata Server in a similar integrated or federated manner.

The federation of metadata involves many issues (see "Metadata Management Technology Integration Cautions and Considerations"), but there are some best practices for minimizing risk and effort while maximizing return on investment (see "Best Practices for Metadata Management"). For example, some less-sophisticated organizations may choose to federate only the production versions of metadata in WSRR with RAM, replacing the metadata artifacts in WSRR with the new versions in RAM as soon as they're ready to go into production. Organizations that can handle the more-complex versioning, administration and analytical issues across repositories should also be able to use the IBM solution in that manner.

Gartner

4.0 IBM's BSR Strategy

RAM and WSRR are key components in IBM's overall BSR strategy. WSRR provides the basis for BSR because of its integration with (and it's a mandatory prerequisite for) the WebSphere Business Services Fabric solution. IBM's business models are based on the World Wide Web Consortium's (W3C's) Web Ontology Language (OWL), which is the standard IBM uses to classify and store metadata. Suitable subsets of this business service metadata (provided as part of BSR functionality with WSRR and Fabric) are planned to be federated into RAM. There will also be cases in which the services will be deployed onto Fabric (executed by WebSphere Process Server), for which metadata will reside solely in WSRR.

5.0 Netting It Out

From an IBM customer's perspective, the metadata management support for the first three communities (system, data and SOA) is an improvement compared with a year ago. IBM can be expected to provide enhancements to these three repositories on an evolutionary basis, based on customer feedback. In addition, IBM must release the RAM repository and gain experience in 2007 and 2008 with best practices in its use. Customers will undoubtedly identify new use cases requiring changes, so we'll have to wait and see how IBM responds to these potentially unforeseen requirements.

Organizations with IBM and other heterogeneous technologies should evaluate each community-of-user repository independently of the others, because they may find that each repository has its own advantages. For example, we strongly advise organizations that are strategically using IBM WebSphere for their SOA solutions to instead consider using WSRR. However, if they aren't a Rational or Tivoli user, then they may not want to use the other IBM repositories, instead opting for others from their development or system technology providers. In addition, we strongly advise organizations that have a breadth of IBM tools in all categories to consider using all four IBM repositories to gain the value-added capabilities provided by federation.

IBM is in the midst of rolling out RAM; therefore, defining exactly how the federation of metadata will be conducted is arguably a bit premature (although IBM has a clearer picture of how RAM and WSRR will be federated to support a BSR strategy than how to federate RAM with the other repositories). In terms of federation, success or failure is in the details. Customers must ask some key questions:

- Which types of artifacts will be supported in each repository?

- Which metadata assets will be redundantly stored in more than one repository, and synchronized?

- What will be the best practices for administering metadata that crosses the repositories?

- Because all metadata integration happens in the "federated metadata layer," which technologies and best practices will be provided to avoid "integration spaghetti"?

IBM will need time to evolve its metadata management strategy and tools to answer these questions, but it already has a wealth of experience and combined strengths to confront the problem — including knowledge and resources involving WebSphere data management and integration, as well as Rational development, life cycle and change management, and configuration management methods and tools. The greatest likelihood is that IBM will evolve the vision forward into a credible deliverable that will meet the needs of mainstream, IBM-centric organizations reasonably well, and for user communities of IBM tools in heterogeneous environments. However, more-sophisticated organizations will need to consider augmenting IBM's tools with other repositories.

Gartner

Customers using IBM products today will likely choose RAM because of its tight integrations with Rational ClearCase, Rational ClearQuest and WSRR. Because the RAM repository is customizable ("extensible"), in the long term it may be able to provide comparable functionality to leading metadata repositories. However, customers needing more-robust legacy metadata management, or metadata warehousing capabilities through at least 2010, should consider augmenting the IBM repositories with one from other leaders, such as Allen Systems Group (Rochade and becubic), BEA (AquaLogic Enterprise Repository) and LogicLibrary (Logidex).

RECOMMENDED READING

"The Evolving Metadata Repository Market"

"Metadata Management Technology Integration Cautions and Considerations"

"Best Practices for Metadata Management"

Acronym Key and Glossary Terms

BPM	business process management
BSR	business service repository
CCMDB	Change and Configuration Management Database (Tivoli)
ESB	enterprise service bus
OWL	Web Ontology Language
RAM	Rational Asset Manager (IBM)
RAS	Reusable Asset Specification (Object Management Group)
SOA	service-oriented architecture
W3C	World Wide Web Consortium
WSMS	WebSphere Metadata Server
WSRR	WebSphere Service Registry and Repository

Gartner

REGIONAL HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
U.S.A.
+1 203 964 0096

European Headquarters
Tamesis
The Glanty
Egham
Surrey, TW20 9AW
UNITED KINGDOM
+44 1784 431611

Asia/Pacific Headquarters
Gartner Australasia Pty. Ltd.
Level 9, 141 Walker Street
North Sydney
New South Wales 2060
AUSTRALIA
+61 2 9459 4600

Japan Headquarters
Gartner Japan Ltd.
Aobadai Hills, 6F
7-7, Aobadai, 4-chome
Meguro-ku, Tokyo 153-0042
JAPAN
+81 3 3481 3670

Latin America Headquarters
Gartner do Brazil
Av. das Nações Unidas, 12551
9° andar—World Trade Center
04578-903—São Paulo SP
BRAZIL
+55 11 3443 1509

Gartner