

Using information as a service to facilitate business-process automation.

Contents

- 2 Introduction
- 3 The Information Server advantage
- 8 WebSphere Information Services Director
- 12 Types of information services
- 13 Information-service consumers
- 16 Creating an information service for business-process automation
- 20 IBM Information Server SCA support
- 22 Conclusion
- 23 For more information

Introduction

Service oriented architecture (SOA) is a business-centric IT architectural approach that supports integrating your business as linked, repeatable business tasks, or services. When you use an SOA to provide information as a service across the enterprise, you can improve the access to and consistency of information. And you can use the data in more ways. Open, industry standards, when employed with information-integration technologies, enable you to decouple the information stored in silos from the technology of those silos and to publish reusable information services. These information services can then be used by processes or people-centric solutions from across diverse operating platforms, applications and existing systems. With the right information services in place, decision makers at all levels in your organization can have timely, enriched access to authoritative, personalized and trusted information. By removing the traditional barriers to information sharing, you can drive innovation with deeper and more-accurate insight into operational, transactional, analytical and unstructured information. SOA teaches that business flexibility is achieved by accessing business functions, data and processes as services, which can be hosted on disparate systems and by different service providers - wiring them together in new and creative ways to create competitive advantage. Service-oriented business applications fulfill this vision.

This white paper addresses one of the key techniques for wiring together business functions by combining the capabilities of information services with the world of business-process automation. It is intended to help enterprise architects and other technical decision-makers understand the benefits of combining IBM Information Server with IBM WebSphere® Process Server to create data-driven business processes that can be designed and deployed more quickly, with a higher degree of confidence in the data, while retaining the flexibility of SOA.

The Information Server advantage

IT executives tell IBM that they need to do a better job of using information. They complain about "information complexity," or a "deluge of information," but what that really means is that there is a great deal of valuable information locked away in various databases and systems throughout the business, but there is no easy way of using this information to gain additional insight into their business, to improve their organization, to compete more effectively or to drive innovation. This challenge is compounded by the fact that organizations tell IBM that information inflexibility inhibits their ability to respond quickly to change. They can't take advantage of new opportunities for innovation, and their costs of maintaining IT systems continuously escalate as the business demands change from systems that weren't built for change. The problem is that when each process creates custom access to information, you end up with:

- Inconsistent views of data across processes (one process gets account data from different places than another process)
- Inconsistent application of rules (calculations were performed differently from process to process)
- Multiple points of maintenance for the same logic (a process that is complex, time-consuming and expensive)

SOA helps centralize and standardize the approach to data integration for processes. Within this approach, information is packaged as a service to business processes, so that consistent, manageable information is made available to every process in a standardized way. An information service is an abstract representation of processing that runs on information providers. A deployed service runs on an application server and processes requests from service-client applications. The developer of a service client knows only about the service interface of the encapsulated operations and the service bindings that are available for interacting with the service. As an information-service developer, you select the job, federated query or other information provider that implements an operation. You can group operations in the same information service or design them in separate services.

This approach enables the user to provide information that has:

- Consistent definition and packaging of data from process to process
- Consistent rules applied to the data
- Improved data quality
- Centralized control and maintenance

Delivering information as a service to people, processes and applications enables organizations to fully use their enterprise data and make it available when and where it is needed to improve operational insight, extend competitive advantage and enable innovation.

IBM Information Server is a platform that helps you derive more value from the complex, heterogeneous information spread across your systems. It does for information what the application server did for Web-based business, providing core services that can be called on to handle difficult integration and management tasks in a repeatable and reliable way. It provides a single layer of shared services that feed trusted enterprise information to all the people, processes and applications that need it (see Figure 1).

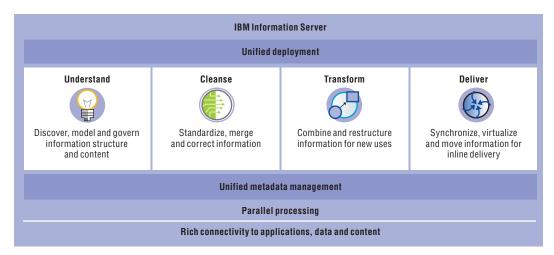


Figure 1. IBM Information Server

IBM Information Server is a fundamental component of any SOA endeavor (see Figure 2). IBM defines three primary entry points for SOA - people, processes and information. Information is delivered as a service using IBM Information Server, helping to ensure that your SOA has enterprise information to use, and that the information is consistent and flexible enough to meet both current and future enterprise needs. IBM Information Server is a revolutionary new software platform that helps organizations derive more value from the complex, heterogeneous information spread across their systems. It enables organizations to integrate disparate data and deliver trusted information wherever and whenever needed, inline and in context, to specific people, applications and processes. IBM Information Server helps business and IT personnel to collaborate to understand the meaning, structure and content of any type of information across any sources. It provides breakthrough productivity and performance for cleansing, transforming and moving this information consistently and securely throughout the enterprise, so it can be accessed and used in new ways to help drive innovation, increase operational efficiency and reduce risk.

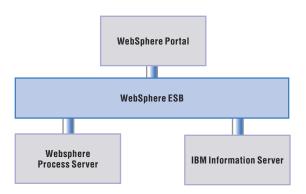


Figure 2. Core components of IBM SOA strategy

IBM Information Server is designed to help companies use their information across all its sources. It delivers all of the functions required to integrate, enrich and deliver information you can trust for your key business initiatives through a set of core product modules: IBM WebSphere Information Analyzer, IBM WebSphere Business Glossary, IBM WebSphere QualityStage, IBM WebSphere DataStage, IBM WebSphere Federation Server, IBM WebSphere Information Services Director and IBM WebSphere Metadata Server. Companion products include IBM Rational® Data Architect and IBM WebSphere Replication Server. IBM Information Server enables you to:

- Understand all sources of information within the business, analyzing their usage, quality and relationships.
- Cleanse it to help assure its quality and consistency.
- Transform it to provide enriched and tailored information.
- Federate it to make it accessible to people, processes and applications.

All of these functions are based on a parallel-processing infrastructure that provides usage and automation across the platform. IBM Information Server also provides connectivity to nearly any data or content source, and the ability to deliver information through a variety of mechanisms.

Underlying these functions is a unified metadata-management foundation that provides seamless sharing of knowledge throughout a project life cycle, along with a detailed understanding of what information means, where it came from, and how it is related to information in other systems.

You can easily deploy and manage integration logic built within IBM Information Server as a shared service within SOA. IBM Information Server provides:

- Access to a broad range of information sources
- A broad range of integration functionality, including federation, extracttransform-load (ETL), inline transformation, replication and event publishing
- High flexibility in how these functions are used, including support for SOA,
 event-driven processing, scheduled batch processing and even standard application
 programming interfaces (APIs) like Structured Query Language (SQL) and Java™

The breadth and flexibility of the platform enable it to address many types of business problems and meet the requirements of many types of projects. This capability helps optimize the opportunities for reuse, which can lead to faster project cycles, better information consistency and stronger information governance.

Delivering information in an SOA context is part of what IBM calls information as a service. This approach to information delivery has many key benefits.

Loose coupling to data stores and data models

An invoking application has no knowledge of how and where the data is stored. As a matter of fact, the service returning the data can be an information service retrieving data from a data store or another application generating data from various sources. This decoupling of application logic from the data model used to store data, such as database schema, helps localize the impact of any future changes in data model only to the implementation of information services. It also facilitates a simple information service to evolve to other types of information services to handle future requirements, such as a change of data sources (as new silos of business applications are integrated), aggregation of data from multiple sources, improving data quality through additional logic and so on.

Reuse of data-access logic

A service-oriented approach embraces separation of data-access logic from the application. This separation can significantly enhance the reusability of the data-access logic.

Support of data governance

It is difficult to track data-access logic interwoven with application logic and dispersed across many applications. Using information services for all data access provides an easily identifiable control point, as well as built-in capabilities for monitoring and fine-grain customizable control of data access. In contrast, without such service orientation it can be extremely complex and time-consuming to determine who is accessing data, how it is being used, or to determine if it is being used in adherence to data governance, security and privacy policies.

Separation of concerns

Using information services separates the development, control and management of data-access logic from development and management of applications and business processes. Application developers and business-process modelers no longer have to possess sophisticated skill sets in data-management technologies. Similarly, database administrators and data owners no longer have to worry about improper access or changing of data by application logic outside of their control.

Ease of development of data services

The data-access service can be easily developed by using high-level tools to specify what data elements should be exposed through each service. And information-management professionals are not required to write detailed Java or C code to invoke SQL calls using database connectors.

Use of existing skill sets and experience

With an SOA approach, application developers do not have to possess database skills. Similarly, information-service developers do not have to possess skills in different programming languages; rather, they use a high-level tool for exposing selected data as a service. In contrast, under direct coupling, application developers need to understand data sources, structure and location.

WebSphere Information Services Director

Publishing consistent, reusable services for information makes it easier for processes to get the information they need from across a heterogeneous landscape. Information integration plays a key role in making this possible. Deployment of integration logic is always done consistently, whether you are using WebSphere DataStage, WebSphere QualityStage or WebSphere Federation Server, using WebSphere Information Services Director. WebSphere Information Services Director enables any of these functions to be easily deployed as Web services, Enterprise JavaBeans (EJBs) or as a Java Messaging Service (JMS), in minutes, without any hand coding. WebSphere Information Services Director load-balances service requests across multiple IBM Information Server nodes, to help ensure smooth pickup of load spikes, and to help ensure fault tolerance and high availability.

The extensible architecture of WebSphere Information Services Director enables it to service-enable a broad range of information-management tasks. Figure 3 shows the components of the WebSphere Information Services Director server.

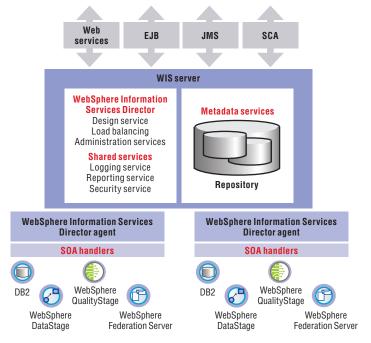


Figure 3. WebSphere Information Services Director architecture

These components include:

Infrastructure services

WebSphere Information Services Director is deployed on a Java 2 Platform, Enterprise Edition (J2EE) technology-based service backbone that provides flexible, distributable and configurable interconnections among the many parts of the architecture. These infrastructure services include:

- Logging services, which provide a central place for a user to record service
 events. Logs go into the common repository with each service provider defining
 relevant logging categories for that service. Configurations determine what
 categories of logging messages are actually saved in the repository.
- Security services, which support role-based authentication of users, access-control services and encryption appropriate for compliance with many privacy and security regulations.

- Service catalog, which provides users with the means to search and browse services by category and to view descriptions available to be defined by WebSphere Information Services Director. (The product also provides a direct link to IBM WebSphere Service Registry and Repository for organizations that have this product.)
- Load-balancing and availability services, which support routing requests to multiple servers to provide optimal loading and a high-availability environment that can recover from any individual server failure.

Operation-provider handlers

As described in the object model, each service is performed by defining operations performed by an operations provider. As shown in Figure 4, WebSphere Information Services Director agents contain handlers to process service requests from the following operations providers:

- WebSphere DataStage, which transforms data of virtually any complexity and delivers it to target applications. WebSphere DataStage provides built-in connectivity for easy access to any source or target system, advanced development tools to define and deploy transformation integration processes and a scalable platform to process massive volumes of data.
- WebSphere QualityStage, which prepares data for integration by providing a
 robust framework for developing and deploying data matching, standardization,
 enrichment and survivorship operations, helping to simplify the process of
 integrating similar data from multiple sources.
- IBM DB2®, which provides a native interface to the flagship IBM relational database system for development and deployment of critical enterprise data.
- WebSphere Federation Server, which presents a single virtual view of the data that can exist in many forms structured and unstructured, mainframe and distributed, and public and private. This data can reside in diverse source systems (such as Oracle databases, enterprise applications, Microsoft® spreadsheets, flat files, the Web, newsgroups and more) and be distributed across a variety of operating environments (such as Microsoft Windows®, Linux®, UNIX® and IBM z/OS® platforms)

Service bindings

Service consumers can access information services using multiple technologies for program interoperability (bindings). WebSphere Information Services Director enables the same service to support multiple protocol bindings, all defined within the Web Services Description Language (WSDL) file. This approach improves the utility of services and therefore increases the likelihood of reuse and adoption across the enterprise. WebSphere Information Services Director provides the unique ability to publish the same service using the following bindings:

- SOAP over HTTP (Web services). Any XML Web service-compliant application
 can invoke an information service as a Web service. These Web services support the
 generation of literal document-style and SOAP-encoded Remote Procedure Call
 (RPC)-style Web services.
- SOAP and Text over JMS. In a message-queue environment, IBM Information Server can automatically generate an asynchronous JMS queue listener (messagedriven bean) and route incoming SOAP messages into information services. As an option, it can adapt the output of an information service into a SOAP message that can be posted to one or more JMS queues or topics.
- EJB. For Java technology-centric development, IBM Information Server can generate a J2EE technology-compliant EJB (stateless session bean) where each information service is instantiated as a separate synchronous EJB method call.
- Service Component Architecture (SCA). This binding provides a client programming model and consistent way of describing components as services available over different protocols. SCA is supported by the IBM enterprise service bus (ESB) product. (See the "Information server support for SCA" section later in this white paper.)

Types of information services

Many services in an SOA environment involve the use of data but not all services have to be an information service. Information services are most often developed in environments where the information infrastructure is sufficiently complex that the use of specialized information-management techniques is required. The following examples show how customers have used IBM Information Server in SOA.

Finance: Federated information-access service

Much of the information within large corporations is stored in a wide variety of data sources from mainframes to highly distributed systems. These | disparate repositories can make it difficult to access information, particularly for new programs. One large European bank encountered this issue as it was attempting to improve its risk-management capabilities to comply with Basel II. The bank had to access information in 23 different retail systems from more than 2500 branches, across many systems. By using federation information-access services based on WebSphere Federation Server, the bank was able to integrate the data of many of its distributed information sources without having to develop a new data warehouse or data store specifically dedicated to risk-bearing information. It now has a real-time view of key risk-bearing data and can access it for regulatory reporting in a highly auditable manner.

Pharmaceuticals: Information-transformation service

A leading pharmaceutical company needed to include real-time data from clinical labs in its research and development reports. The company used WebSphere DataStage to define a transformation process for XML documents from labs. This process used SOA to expose the transformation as a Web service, enabling labs to send data and receive an immediate response. Preclinical data became available to scientific personnel earlier, enabling lab scientists to select which data to analyze. Now, only the best data is chosen, greatly improving scientists' efficiency.

Insurance: Data-cleansing service

An international insurance data-services company uses IBM Information Server to validate and enrich property addresses through Web services. As insurance companies submit lists of addresses for underwriting, services standardize the addresses based on their rules, validate each address, match them to a list of known addresses, and enrich the addresses with more information that helps with underwriting decisions. The company now automates 80 percent of the process and has eliminated most of the errors. All of this was simplified using the SOA capabilities of IBM Information Server and the standardization and matching capabilities of WebSphere QualityStage.

Manufacturing: Service-enabling a data warehouse

A leading manufacturer of automotive machines, parts and accessories needed to reduce the number of out-of-stock and overstock situations across regional distribution centers to more effectively support its network of independent dealerships. By embedding information services into the dealers' inventory-ordering application, the dealers now have a direct integration path to the customer data warehouse that enables them to more accurately predict demand for parts and accessories based on past purchase histories, buying patterns and product availability. The company has now been able to streamline its inventory-management process and reduce its out-of-stock situations by delivering insightful information inline and in context to their dealers. These information services were created using the service-enabling capabilities of WebSphere Information Services Director and the data-transformation capabilities of WebSphere DataStage.

Information-service consumers

To support the usual application interfaces for service delivery, WebSphere Information Services Director provides integrated interfaces to many other critical SOA components. As shown in Figure 4, WebSphere Information Services Director provides interfaces that enable direct access to information services from WebSphere Process Server, IBM WebSphere Enterprise Service Bus (WebSphere ESB), IBM WebSphere Portal and WebSphere Services Registry and Repository.

WebSphere Information Services Director is also a J2EE standard-based application delivered on IBM WebSphere Application Server as its deployment environment.

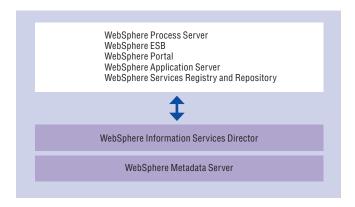


Figure 4. WebSphere Information Services Director interfaces to other SOA components

A key enabler for designing new processes is IBM WebSphere Integration Developer. Based on Eclipse technology, it is an easy-to-learn toolkit for assembling composite applications. As shown in Figure 5, IBM Information Server provides extensions to WebSphere Integration Developer and WebSphere Process Server to use information-service capabilities at design time and run time of automated business processes.

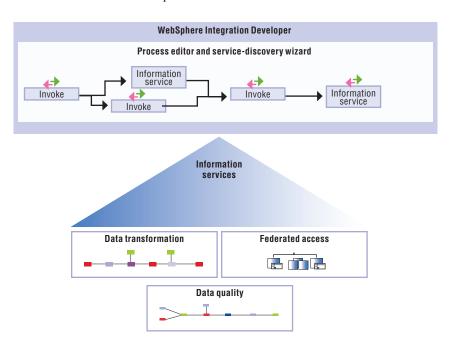


Figure 5. WebSphere Integration Developer provides a design-time toolkit for assembling services.

At design time, it is important to discover information services and understand their values. For that purpose, the process editor and the enterprise service-discovery wizard of WebSphere Integration Developer are extended to allow for browsing information services on an information server, displaying their metadata and importing them into an SOA scenario. The process editor enables you to graphically model workflows using different types of activities. Its extension introduces a new activity type to directly access information services and support SQL.

The enterprise service-discovery wizard provides the ability to discover services from different providers. The extension introduces an adapter to IBM Information Server to enable the direct discovery of information services on IBM Information Server and to import them. The design-time extensions to WebSphere Integration Developer are included as Eclipse plug-ins and can be easily installed using IBM Rational Product Updater that is delivered with WebSphere Integration Developer.

WebSphere Integration Developer enables you to directly deploy a business process to WebSphere Process Server. Alternatively, you can export the process as an enterprise application and install it using the WebSphere Process Server administration console.

Figure 6 shows how at run time, the WebSphere Process Server process-container extension enables direct calls to the WebSphere Information Services Director, and thus handles the invocation of information services as the process runs. The extension supports the WebSphere Process Server event infrastructure with information service-specific events that can be of interest for monitoring purposes, for example.

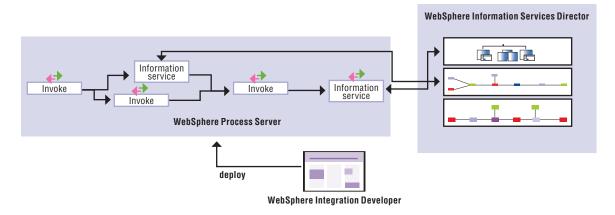


Figure 6. At run time, process flows invoke information services.

Creating an information service for business-process automation

As a developer of business processes using WebSphere Process Server, WebSphere Information Services Director offers a transparent interface to using information services through the WebSphere Integration Developer plug-in that gives you:

- Direct access to information services from within the Business Process Execution Language (BPEL) editor
- Direct support of full SQL from within the BPEL editor
- Direct mapping of in and out parameters to BPEL variables
- Information-service-specific events

This example shows how a business process calls an information service that is implemented by service-enabling a WebSphere DataStage job. WebSphere DataStage is a key information provider that delivers deep information-integration capabilities with support for everything from individual stages, to connections, to entire data flows that can be reused across different jobs and projects. In addition, like the other product modules, WebSphere DataStage uses the shared platform services for parallel processing, administration, deployment and connectivity of IBM Information Server. A service-ready data-integration job accepts requests from client applications, mapping request data to input rows and passing them to the underlying jobs. A job instance can include database lookups, transformations, data standardization and matching, and other data-integration tasks offered by IBM Information Server.

Figure 7 shows the WebSphere DataStage design canvas in which a job has been built that combines data from a company's European and North American operations and joins the orders into a single database table, and then creates subsequent lists of orders that need to be approved and orders that need to processed. If the process receives detected error messages, a human task (a participating task) is created to handle the detected errors. The business process returns a task response or the process input if no task was created.

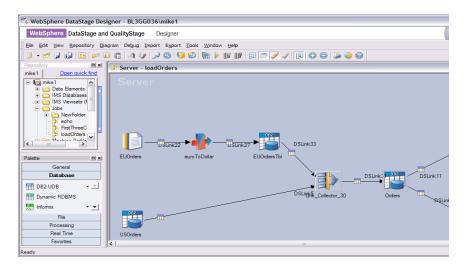
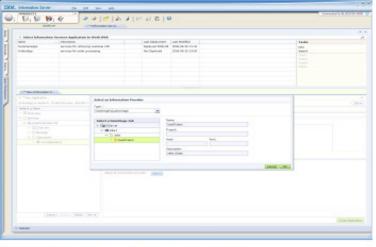


Figure 7. WebSphere DataStage job load and processes for customer-order data

A service operation identifies one or more information providers that perform the work and return results or acknowledgments to a service client. As an information-services designer, you decide which information provider implements an operation. This loose coupling enables service clients to be developed independent of information services, and requires minimal coordination between the service developer and the service-client developer. In this example, the WebSphere DataStage job for loading these European and North American orders is defined as an information service. Depending on its design, a job can accept input parameters from a service client, which guides processing in the job, and return output to the service client. When the information service is saved, the design-time information is saved in the metadata server and can be exported to other servers. Figure 8 shows the wizard-like interface that leads the users through the definition of the "loadOrders" job as an information service. After being defined and saved, the service can then be deployed as a J2EE application.



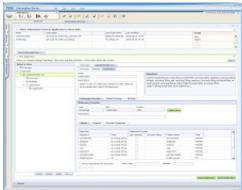


Figure 8. The IBM Information Server console creates and deploys the information service.

After it is deployed, the information service can be invoked through any of its multiple supported bindings (such as Web services, EJB and JMS). For WebSphere Process Server applications, a special interface has been developed to enable you to use information services as part of orchestrations defined in WebSphere Integration Developer. By invoking an information service, you can use the information in a business process in various ways, such as:

- To guide the control flow. For example, you might want to use inventory information to decide if a supplier service needs to be called for items that will shortly be out of stock.
- $\bullet \quad \textit{To retrieve business data and send it to partners, clients or other services}.$
- To update business data. For example, the delivery time of an item can be changed based on the information received from the supplier of this item.

In this example, the WebSphere Integration Developer user can define a business-integration module in which the information service delivers data to the process. The first step is to insert a new information-service activity into the business process. You can then discover an appropriate information service to bind to the new activity by first identifying the WebSphere Information Services Director server in which the services are cataloged and deployed.

WebSphere Integration Developer enables you to directly browse the contents of the metadata server in IBM Information Server to obtain details about available services, so you can discover deployed services and operations (see Figure 9).

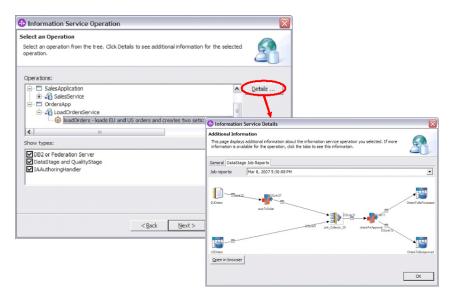


Figure 9. The browse function provides detailed metadata about the information service.

As shown in Figure 10, after you have selected the service, you can view additional detailed information about the selected operation, including implementation details. The input/output (I/O) parameters for the selected information service are then mapped to business-process variables. Complex data types that are imported through the operation selection are automatically added to the list of selectable data types.

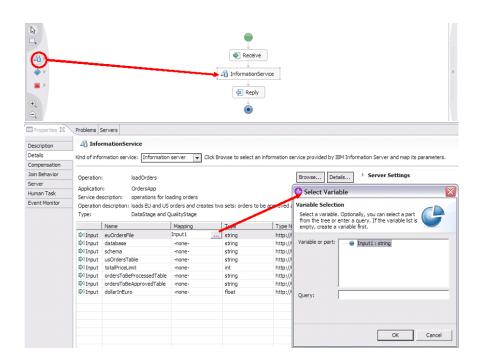


Figure 10. Mapping parameters to the business process variables

After the business module containing the information service has been added to the WebSphere Process Server configuration, information-server runtime support within WebSphere Process Server enables it to invoke the new process.

IBM Information Server SCA support

SCA is a set of specifications that describe a model for building applications and systems using SOA. It extends and complements prior approaches to implementing services, and builds on open standards such as Web services. SCA makes it possible to use SOA to organize business application code based on components that implement business logic. These components offer their capabilities through service-oriented interfaces and consume functions offered by other components through service-oriented interfaces, called service references.

WebSphere ESB uses SCA to define its interfaces to its mediation flows. The enterprise service-discovery wizard is a tool that guides the ESB developer through the discovery of existing services. WebSphere Information Services Director provides an extension to the enterprise service-discovery wizard that enables you to browse WebSphere Information Services Director service catalog. This extension shows some of the specialized metadata from the information service, including operational metadata. Based on this metadata, you can then decide which service fits best. For each operation to be imported, an import binding based on Web services is created in the assembly editor.

Discovering information services using the enterprise service-discovery wizard requires the following steps:

- Open the enterprise service-discovery wizard and select the WebSphere Information Services Director service provider.
- Connect to a WebSphere Information Services Director instance by providing server name, ports, user ID and password.
- Browse a tree of WebSphere Information Services Director applications, projects, services, operations and operation parameters (shown in Figure 11). For operations, you use the details button in the enterprise service-discovery wizard to show metadata that contain, among other things, service description, date created, WebSphere DataStage job reports (if applicable) and a federated query of an information-integration service (if applicable). It is also possible to filter according to operation types (such as WebSphere DataStage or WebSphere Federation Server).
- Select operations to import, optionally selecting for which subset of bindings the import should be performed.
- Select the SCA module where the service should be imported.

After all artifacts have been created, you can use the assembly editor in the usual way to integrate the new Web services-based import binding in the SCA module.

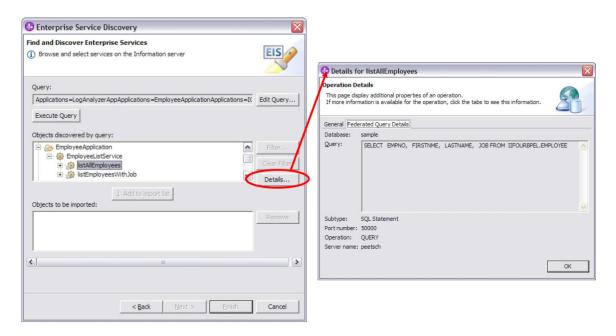


Figure 11. The enterprise service-discovery wizard shows metadata for operation types.

Conclusion

Overall, IBM Information Server provides significant advantages to developers of business-process applications and dealing with complex, heterogeneous information. These advantages include:

- A comprehensive, unified foundation for enterprise information architectures, scalable to practically any volume and processing requirement
- · Auditable data quality as a foundation for trusted information across the enterprise
- Metadata-driven integration, providing breakthrough productivity and flexibility for integrating and enriching information
- Consistent, reusable information services along with application services and process services, an enterprise essential
- Accelerated time to value with proven, industry-aligned solutions and expertise
- Broad and deep connectivity to information across diverse sources, such as structured, unstructured, mainframe and applications

Specific to delivering information as a service, IBM Information Server now has the capabilities to help ensure that information can be efficiently used across the enterprise to drive innovation, increase operational insight and improve the flexibility of organizations to compete more effectively.

To maximize the benefits of SOA, business-process automation cannot be considered separately from information integration—the two are mutually dependent for success in supporting the dynamic nature of business today. By using IBM Information Server and IBM WebSphere Process Server to decouple information from its individual data sources and make it easily consumable by business processes and applications, you can help ensure that the best information is delivered when and where it is needed, as your business needs dictate.

For more information

To learn more about delivering information as a service, contact your IBM representative or visit:

ibm.com/software/data/information-on-demand/soa.html

To learn more about IBM Information Server integration, contact your IBM representative or visit:

publib.boulder.ibm.com/bpcsamp/index.html?productIntegration&productIntegration/infoServer.html

To view the online documentation available at the information center of WebSphere Integration Developer, Version 6.0.2, visit:

publib.boulder.ibm.com/infocenter/dmndhelp/v6rxmx/index.jsp?topic=/com.ibm.is.bpel.help.doc/topics/accessdata.html



© Copyright IBM Corporation 2007

IBM Corporation Software Group Route 100 Somers, NY 10589 U.S.A.

Produced in the United States of America 05-07

All Rights Reserved

DB2, IBM, the IBM logo, Rational, WebSphere and z/OS are trademarks of International Business Machines Corporation in the United States, other countries or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries or both.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries or both.

UNIX is a registered trademark of The Open Group in United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

Other company, product and service names may be trademarks or service marks of others.