

An Overview of issues for the interoperability of information systems

Christophe Nicolle, Jean-Claude Simon, Kokou Yétongnon
Laboratoire L2I, Université de Bourgogne
BP 47870, 21078 Dijon Cedex, FRANCE

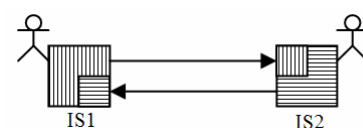
INTEROPERABILITY ARCHITECTURE OVERVIEW

An information system is a multilevel system characterized by: a “Data” level, a “Behavioral” level and a “Communication” level. The data level represents the data stored by the system. The behavioral level represents management and production processes carried out by the system. The processes can interact with the data level to extract, generate and store data. The communication level relates to the network used to exchange data and activate processes between geographically distant users or machines.

Information system interoperation has emerged as a central design issue in web based information systems to allow data and service sharing among heterogeneous systems. Data heterogeneity stemming from the diversity of data formats or models used to represent and store information in the web is a major obstacle to information systems interoperability. These data models range from the structured data models (network, relational, OO) found in traditional databases to flat files and emerging web oriented semi-structured models. Information system interoperability aims at supporting the amalgamation autonomous heterogeneous systems to create integrated virtual environments or architectures in which information from multiple disparate sources can be accessed in a transparent and efficient manner. As an example of such integrated virtual systems, consider an airline reservation system based on the integration of a group of airlines reservation and ticket sale information systems. The specific airline systems provide various types of fares and special discount trips. That can be searched and compared to respond to user queries for finding the best available prices for specified flights.

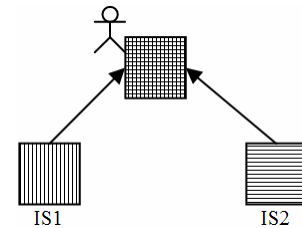
Database interoperability issues have been extensively studied in the past. Several approaches, including database translation, distributed systems, federations, language based multidatabase, ontology and mediation, have been proposed to bridge the semantic gaps among heterogeneous information systems.

The **database translation** approach is a point-to-point solution based on direct data mappings between pairs of information systems. The mappings are used to resolve data discrepancies among the systems (Yan L.L., Ling T.W., 1992). The database translation approach is most appropriate for small scale information processing

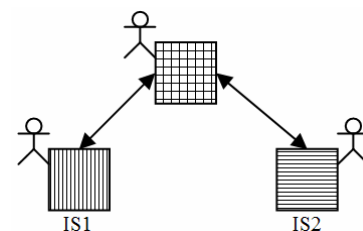


environment with a reduced number of participants. The number of translators grows with the square of the number of components in the integrated system. For example, consider two information systems IS1 and IS2 in the travel agency example above, the corresponding translators must be placed between the information systems as shown in the figure above. Information in IS1 is represented by vertical lines while the information in IS2 is shown as horizontal lines.

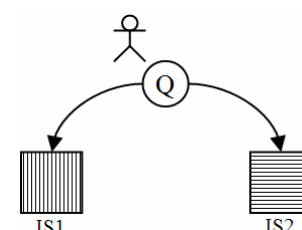
In the **standardization** approach, the information sources use the same model or standard for data representation and communication. The standard model can be a comprehensive metamodel capable of integrating the requirements of the models of the different components (Atzeni P., Torlone R., 1997). The use of a standard metamodel reduces the number of translators (this number grows linearly with the number of components) to resolve semantic differences. However, the construction of a comprehensive metamodel is difficult; the manipulation of high level languages is complex; and there is no unified database interfaces. In our example, the travel agencies must define a common model to export their data. A centralized information system can be built to replace the original information systems (IS1, IS2). The global centralized schema is a combination of the entire data (horizontal and vertical lines) contained in IS1 and IS2.



Federated systems consist of a set of heterogeneous databases in which federation users can access and manipulate data transparently without knowledge of the data location (Sheth A.P., Larson J.A., 1990). Each federation database includes a federated schema that incorporates the data exported by one or more remote information systems. There are two types of federations. A tightly coupled federation is based on a global federated schema that combines all participant schemas. The federated schema is constructed and maintained by the federation administrator. A loosely couple federation includes one or more federated schema that are created by users or the local database administrator. The federated schema incorporates a subset of the schema available in the federation. This approach becomes rapidly complex when of number of translators required becomes large. In our example, the existing information systems are completely operational for local users. Only the shared data are integrated in the federated schema. The federated system is only made of horizontal and vertical lines that IS1 and IS2 want to exchange.

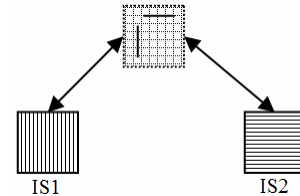


Language based **multi-base systems** consist of a loosely connected collection of databases in which a common query language is used to access the contents of the local and remote databases (Keim D.A., Kriegel H.P., Miethsam A., 1994). In this approach, in contrast to the distributed and federated systems, the burden of creating



the federated schema is placed on the users who must discover and understand the semantics of the remote databases. In our example, the various companies have to define a global common language (Q) to query their information systems (IS1, IS2). This solution is well adapted for information systems that are based on the same family of data models and do not require complex query translators.

The **ontology** based interoperability approach uses ontology to provide an explicit conceptualization of the common domain of a collection of information systems (Benslimane D., Leclercq E., Savonnet M., Terrasse M.N., Yétongnon K., 2000). An ontology defines a common vocabulary that can be used by users from different systems. The construction an ontology for a domain is a difficult task and often requires merging existing overlapping ontologies. The interoperability solutions based on ontology describe the semantics of information rather than their organization or their format. In our example, the companies have to define ontology to capture the semantics of their domain of activity.



The **mediation** approach is based on two main components: mediator and wrapper. The mediator is used to create and support an integrated view of data over multiple sources. It provides various services to support query processing. For instance a mediator can cooperate with other mediators to decompose a query into sub queries and generates an execution plan based on the resources of the cooperating sites. The wrapper is used to map the local databases into a common federation data model. The wrapper component provides the basic data access functions (Garcia-Molina H., Hammer J., Ireland K., Papakonstantinou Y., Ullman J., Widow J., 1995). In our example, a translator, which acts as a wrapper, is placed between the conceptual representation of the mediator and the local description of each information source

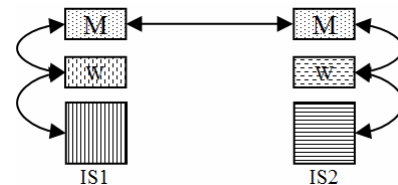


Table 1 summarizes the various architectures for the interoperation of information systems. In this table, a brief presentation of the advantages and limits of each approach is given.

Table 1 Overview of architectures for interoperable information systems

D: Data level, B: Behavioral level, C: Communication level.

Systems	Advantage	Limits	Tools or methods used	Level
Translation	Better control of point to point translation.	- Requires a large number of translators in open environments. - Adding a new Information System requires $2(n-1)$ translators.	Required $n*(n-1)$ translators	☒ D ☐ B ☐ C
Standardization	- Use of pivot, canonical model or metamodel. - Reduce the number of translators.	- Definition of a common standard accepted by all IS. - The construction of a comprehensive metamodel is difficult.	Required $2n$ translators	☒ D ☐ B ☐ C
Federation	Derived from	Use of a global, static federated	Required $2n$	☒ D ☐ B

	standardization. • Local IS are autonomous.	schema. • The construction of an integrated federal schema is difficult. • New addition requires redesign of federated schema.	translators	☐ C
Multi-base	• Used of a single language for many IS.	• The common interoperate language does not export local system semantics. • Users need to discover and understand the semantic of remote IS.	Query Based	☒ D ☒ B ☐ C
Ontology	• Semantic oriented solution.	• Extensive ontologies are voluminous. • Requires meta-level translation.	Semantic	☒ D ☐ B ☐ C
Mediation	• Combine translation and semantic. • Local IS are autonomous.	• Difficult to construct automatic mediator process.	Required 2n semantic translators	☒ D ☒ B ☐ C

EMERGING INTEROPERABILITY APPROACH

As new data models are developed for web based information systems, there is a need to extend interoperability solutions to take into account requirements and specifications of the new models. For instance, **XML** (XML, 2004) has emerged recently as an important model for describing and sharing web based data. This importance stems from two major factors. First, XML is becoming a de facto data standard supported by many software vendors and applications developers. Second, XML is based on a relatively simple structure that is both user and machine readable and that can be used by non expert database administrators. The existing web technologies are not initially intended to address some of the issues involved in database integration. For instance the web-browsing paradigm is efficient for data look up in large environment, but is inadequate for database integration support. To use this paradigm to locate and merge data requires costly applications that are often tailored to specific integration needs. New challenges have arisen from the development of web based information systems. One of the challenges is the need to develop web-oriented tool to support information integration and allow access to local as well as remote information sources.

Recently, web services have been proposed as a method to address some of the challenges of web-based integrated systems. A web service can be viewed as a set of layers contained in a stack. The layers are dynamically defined following user needs and are called through a set of Internet protocols. The protocols are different those propose for various network architectures. However, in all web service architecture, a base set of protocols is always used (W3C, 2002). This base set is composed of: SOAP (SOAP, 2003), WSDL (WSDL, 2003) and UDDI (UDDI, 2002). They allow the discovery, description and information exchanges between web services.

SOAP is a mechanism that uses XML for the exchange of structured and typed information between several actors in a decentralized and distributed environment. SOAP does not define the semantics of the application, but provides a mechanism for expressing semantics by proposing a modular template and mechanisms for data coding.

WSDL uses XML syntax to describe the methods and parameters of web services. These parameters include: protocols, servers, ports, input and output messages format, and exceptions format. With WSDL, an application using SOAP can auto-configure the web services exchanges, masking the majority of the low-level technical details.

UDDI is a web-based company world directory combining "white pages" (Information such as name, address, telephone number, and other contact information of a given business), "yellow pages" (Information that categorizes businesses) and "green pages" (Technical information about the web services provided by a given business.). UDDI allows web service references by automating all search procedures. Table 2 presents the advantages and limits of web services.

In our example, a set of web services can be built from each information system independently from the other information systems. The web services become a standard interface to access the local information system. These web services can be used by customers and partners via the internet and by local users via an intranet. This solution is very flexible and reduces the complexity of the heterogeneity problem.

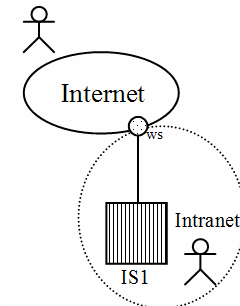


Table 2. Web services, new architecture for interoperability.

Systems	Advantage	Limits	Tools or Methods used	Level
Web services	- Resolved format level. - Resolved process translation. - All level of IS are managed. - Normalized solution. - Developed both by industrials and researchers.	- Security mechanism not finalized. - Combination of web services not resolved.	Protocol SOAP, WSDL, UDDI	☒ D ☒ B ☒ C

To achieve web service architecture, several industrial tools have been developed. Four main actors in the industrial world share the market. The solutions proposed by Microsoft and SUN are language oriented while the solutions proposed by IBM and BEA are platform oriented.

- **Microsoft .NET** proposes a software platform on which companies can exchange data and services on the Internet based on an ASP model (Application Provider Service). Most Microsoft products can be extended to use web services developed with the .NET. The philosophy of this solution can be resumed by "one OS, many languages".
- The **SUN J2EE** is developed by the Java Community Process. It is a set of services and specification containing JDBC (Java Data Base Connector), JMS (Java Message Services), JSP (Java Server Pages), EJB (Enterprise Java Beans) etc. J2EE 1.4 includes web service specifications using an open source framework called AXIS (used by **IBM WebSphere**). In response to the **Microsoft .NET** solution, SUN proposes **ONE** which groups the set of SUN web services propositions. The philosophy of this solution can be resumed by "many OS, one language".
- **IBM WebSphere** is a set of components allowing the creation of interoperable information systems based on web services. These components include Interchange Server which allows process integration; MQ Integrator Broker which allows data integration; MQ Workflow

which allows processes management etc. **IBM WebSphere** uses the SUN JAVA language for the development of its web services.

- The **BEA WebLogic Server** is based on the Java Connectors Architecture. This tool uses the notion of components and connectors which can be integrated between them. The integration of the connectors is carried out by the Application Integration framework and the Adapter Development Kit. To manage the resulting architecture, a Business Process Management is used in coordination with the B2B integration tool. This tool exploits standards such as XML, HTTP or SSL and semantic solutions such as RosettaNet, cXML, ebXML, EDI.

C O N C L U S I O N

For the past 20 years or so, the need to exchange information between various partners pushed researchers to develop architectures for the interoperability of information systems. The proposed architectures have addressed several key interoperability issues, ranging from the resolution of data format heterogeneity using translations based architecture and the reduction of the number of required translators in standardization based architecture to the resolution of semantic heterogeneity based on ontology and the resolution of process heterogeneity with mediation based architecture.

Nowadays, information systems can be integrated or disassociated depending on the market trends of enterprise mergers. The web service based architecture allows the development of this type of interoperability by proposing: a standard data format with XML, a standard communication architecture based on the SOAP protocol and a standard description of processes using WSDL and UDDI. The next major challenge in the web service world is to extend web services to include security, data owner and semantics.

R E F E R E N C E S

Atzeni P., Torlone R. (1997). MDM: A Multiple-data-Model tool for the management of Heterogeneous Database Schemes, Proceedings of the SIGMOD International Conference, pp 538-531.

Benslimane D., Leclercq E., Savonnet M., Terrasse M.N, Yétongnon K (2000): On the definition of generic multi-layered ontologies for urban applications, International Journal of Computers, Environment and Urban Systems, Volume 24 (2000), pp 191-214, Elsevier Science Ltd, England.

Garcia-Molina H., Hammer J., Ireland K., Papakonstantinou Y., Ullman J., Widow J. (1995) Integrating and Accessing Heterogeneous Information Sources in TSIMMIS, Proceeding of the AAAI Symposium on Information Gathering, pp. 61-64, Stanford, March 1995.

Keim D.A., Kriegel H.P., Miethsam A. (1994) Query Translation Supporting the Migration of Legacy Database into Cooperative Information Systems, Proceedings of the 2nd International Conference on Cooperative Information Systems, pp203-214, 1994

Sheth A.P., Larson J.A.(1990) Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases, ACM computing surveys, vol. 22, No 3, 1990.

SOAP (2003) SOAP Version 1.2 Part 0: Primer, W3C Recommendation 24 June 2003, <http://www.w3.org/TR/soap12-part0/>

UDDI (2002) UDDI Version 3.0, UDDI Spec Technical Committee Specification, 19 July 2002, http://uddi.org/pubs/uddi_v3.htm

W3C (2002), Web Services Architecture Requirements, W3C, Working Draft 14 November 2002, <http://www.w3.org/TR/2002/WD-wsa-reqs-20021114>

WSDL (2003) Web Services Description Language (WSDL) Version 1.2 Part 1: Core Language W3C Working Draft 11 June 2003, <http://www.w3.org/TR/wsdl12>

XML (2004) Extensible Markup Language 1.0 (Third Edition), W3C Recommendation 04 February 2004, <http://www.w3.org/TR/REC-xml>

Yan L.L., Ling T.W. (1992): Translating Relational Schema with Constraints into OODB Schema, Proceedings of the IFIP WG2.6 Database Semantic Conference on Interoperable Database Systems (DS-5), Lorne, Victoria, Australia, pp 69-85, 16-20 November, 1992.

Terms and Definitions

Interoperability: The ability of heterogeneous software and hardware to communicate and share information.

Ontology: An explicit formal specification of how to represent the objects, concepts and entities existing in some area of interest and the relationships among them.

SOAP (Simple Object Access Protocol): A XML-based message protocol used to encode information in web service requests and response messages before sending them over a network. SOAP messages are independent of any operating system or protocol and may be transported using Internet protocols (SMTP, MIME, and HTTP).

UDDI (Universal Description, Discovery and Integration): A web-based distributed directory for discovery of web services offered by companies. It is similar to a traditional phone book's yellow and white pages.

Web service: A software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-readable format (specifically WSDL). Other systems interact with the web service in a manner prescribed by its description using SOAP-messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards.

WSDL (Web Services Language Description): An XML-formatted language used to describe a web service's capabilities as collections of communication endpoints capable of exchanging messages.

XML: A language for creating markup languages. There are two kinds of XML documents: well-formed and valid. The first respects the XML standard for the inclusion and the names of the tags. The second must well-be formed and uses a grammar to define the structure and the types of the data described by the document.