An Approach for the Integration of Data within Complex Logistics Systems

Carl Hans, Karl A. Hribernik, Klaus-Dieter Thoben
BIBA / University of Bremen
{han, hri, tho}@biba.uni-bremen.de

Abstract

Due to a global competition today's business world changed dramatically whereas the ongoing trend towards global enterprise networks has lead to increasing customer expectations regarding the performance of logistics systems which have to be simultaneously reliable, robust and cost-effective. One the one hand reliability, robustness and cost-efficiency mainly depend on the availability of suitable ICT-solutions as well as their proper integration into the business processes. As a consequence various systems were developed in the past and are currently in use. On the other hand the seamless access to the data kept by all these different systems is a key issue in order to meet the requirements of logistics systems of today and tomorrow. In this context data integration represents a major challenge which is still unsolved.

This paper describes a proposal for a service-oriented approach towards the seamless integration of logistics data which aims at combining existing systems and standards and thus overcoming today's data and information barriers. By accepting existing standards and interfaces this requires innovative approaches addressing the semantic description of data, their transparent transformation between different representations as well as a common interface which can be provided to the various systems which are currently applied in the field of logistics.

Motivation

The evolution of today's business towards enterprise networks requires logistics systems which are simultaneously reliable, robust and cost-effective. In order to fulfil these needs the application of new technologies (e.g. RFID) is an absolute must. Latest initiatives in this field go beyond that by evaluating and developing a new paradigm of autonomous, cooperating logistics processes (http://www.sfb637.uni-bremen.de). Alongside the increased integration of ICT into logistics processes the seamless access to accompanying logistics data became a major issue. A general problem in this field results from the existence of various data sources and accompanying data management and support systems as well as a diversity of underlying standards and protocols.

For competitive reasons, there is usually no interest in supporting a general interface for data access from the perspective of solution providers. Additionally, as end users, logistics service providers are interested in keeping their data safe, secure and protected in order to eliminate the potential risk that their data can be accessed, interpreted and applied by their competitors to improve their process performance. Thus the huge variety of data sources is treated in an isolated manner and the potential for a further improvement of logistics processes remains unused.

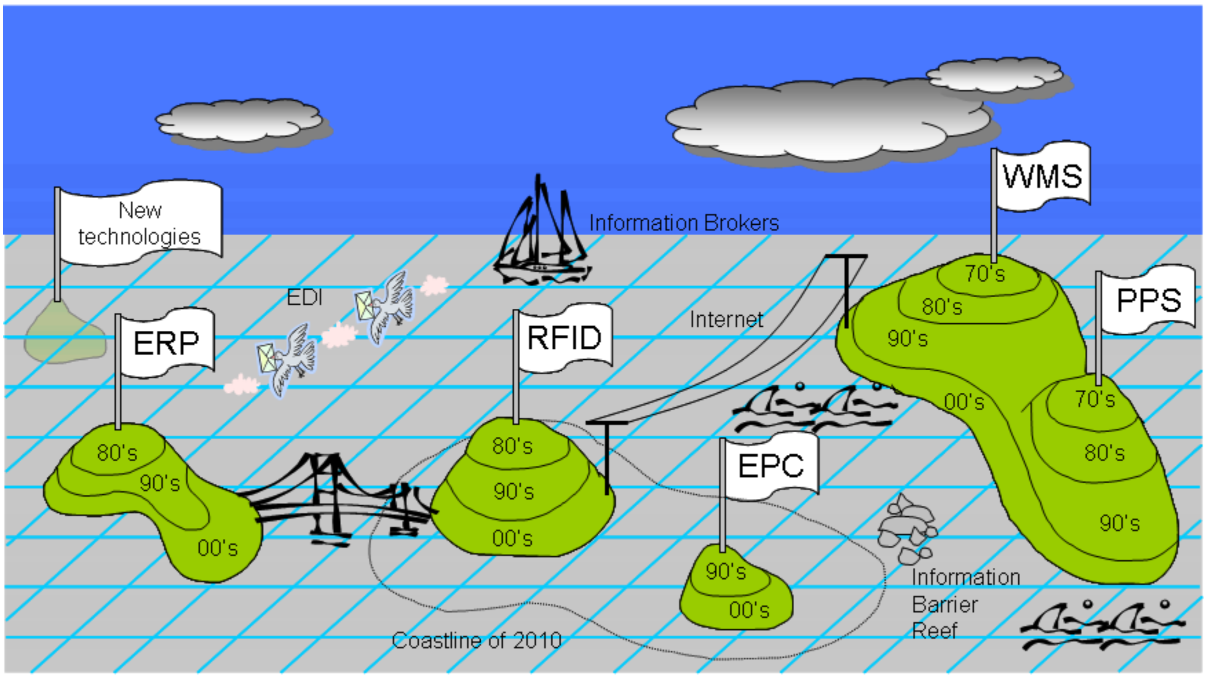

Figure 1: Bridging Islands (based on Hannus 1996)

As shown in Figure 1 significant effort was and still is spent in order to achieve at least a coupling between systems of certain business partners by bridging the technological islands through specific ICT solutions. However most of these solutions will become obsolete sooner or later due to the ongoing development of the islands which is reflected by a continuously decreasing sea level as well as the highly dynamic partnerships within today's enterprise networks. Instead of developing solutions for 1:1 relationships a general solution must be found which allows a unique access to all relevant logistics data while accepting the diversity of existing systems and standards.

In addition to the physical access to data such an approach has to address their semantics towards a general, system-independent and seamless integration of logistics data in order to allow the interpretation and exchange of data between the various systems of the different partners involved in logistics.

A real integration therefore requires a general approach in order to describe data semantics as well as transformation approaches which do the conversion of data into a format the recipient will understand.

Challenge

The requirements set by complex logistics systems towards the integration of data regarding the individual entities within them prove immensely challenging. In order to implement complex behaviour with regards to autonomous control, dynamism, reactivity and mobility, these entities, including objects such as cargo, transit equipment and transportation systems but also software systems such as disposition, Enterprise Resource Planning (ERP) or Warehouse Management Systems (WMS) require the development of innovative concepts for the description of and access to data.

The availability of context and environmental data regarding logistic entities is especially relevant to the field of autonomous control. Here, real-time, local decision-making takes the place of central planning and control. It is thus imperative for the individual logistic entity to be able to access, understand, process and thus integrate context data stemming from a host of diverse, mobile and sometimes unreliable data sources, such as RFID readers, sensor networks, embedded systems, etc. This scenario sets the problem at hand apart from traditional data integration questions which generally deal with static, predictable and reliable data sources such as enterprise or database systems.

In this context, heterogeneous environments such as those found in complex logistics systems demand a unique perspective be taken upon the representation and meaning of data. Here, the heterogeneity and distribution of the syntactic and semantic representations and descriptions of the individual entities within such systems pose the greatest challenge. However, in order for such a complex logistics system to operate, the individual system components need to be able to access and understand data regarding arbitrary entities in a predictable and reliable manner, regardless of the formal, structural and physical properties of the underlying, individual systems.

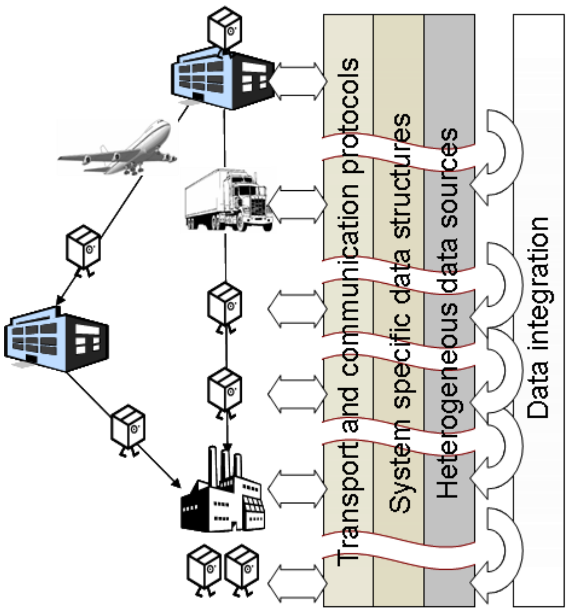

Figure 2: Levels of Required Data Integration

Figure 2 above summarises the challenge towards data integration for complex logistic systems, especially those facilitating autonomous control. The data integration approach must be capable of bridging the semantic and syntactic gaps in the digital representation of arbitrary logistics entities caused by the storage of data across distributed heterogeneous data sources and system specific data structures. Furthermore, it must take into account the restrictions and requirements imposed by the necessity for communication across different transport and communication protocols in complex logistics systems.

State of the Art

Currently, a plethora of systems, data representation models and exchange formats are in use in logistics systems in industry. The logistics system landscape is diverse and complex. Solutions such as Global Connect, UPS Online Tools, PAS and EasySped are employed by logistics service providers and are each based on specific proprietary data models. The same can be said for the ERP systems in use in the field of logistics, which themselves strive to implement the flexibility to represent processes covering a wide range of sectors.

The main standardising factors to be found in this landscape are predominantly common data exchange formats such as specific EDIFACT subsets (Ballnus, 2000) such as FORTRAS or the interface languages of popular ERP systems. Furthermore, XML-based formats play an increasing role due to their ability to represent data in a platform independent, structured way (Yergau et al, 2004).

For example, the EPC (Electronic Product Code) Network offers the EPCIS (Electronic Product Code Information Services) specification, which concentrates on defining an interface to views upon item-specific product data within supply chains. EPCIS can successfully be employed to model lifecycle dynamics and other related areas, and can readily be extended in scope, syntax and semantics by means of "user vocabularies". However, the strong focus on item-specific tracking and tracing remains dominant in the scope of application of the EPCIS standard.

The Integrated Project PROMISE (Product Lifecycle Management and Information Tracking Using Smart Embedded Systems) is currently developing a standardised approach to exchange instance-specific product information in an XML messaging format, PMI (Promise Messaging Interface). As the name implies, this format focuses on providing access to product information and concentrates on adequately representing the product lifecycle. Whilst significant areas of lifecycle management do overlap with those of logistics, PMI doesn't explicitly support item-specific logistics information exchange.

Solutions such as Seeburger Business Integration Server do indeed provide the technology to integrate heterogeneous enterprise systems across all stakeholders of a supply chain. However, the challenges of data integration in complex logistics systems go beyond the focus of such solutions, which mainly concentrate on the optimisation of trade relationships by integrating large-scale enterprise systems such as CRM, SCM and ERP. For one, the type of item-specific granularity with regards to logistics entities are outside of the scope of such solutions, for another, the integration of real-time, dynamic data sources such as sensor network or embedded systems are generally not part of this class of integration server.

Approach to Data Integration

The diversity of systems involved in complex logistic processes defines challenging requirements towards an approach to data integration in this field. The variety of different applications, data sources, exchange formats and transport protocols demand a high level of flexibility and scalability.

To address these challenges, the literature in the field of traditional data integration needs to be studied bearing in mind the specific requirements complex logistics systems exhibit. Here, a number of different, traditional solutions to data integration may be taken into account, foremost amongst these tightly coupled, loosely coupled and object-oriented approaches (Wache 2003). Whilst a tightly coupled approach can quickly be dismissed on grounds of its inflexibility, loosely coupled and object-oriented approaches cannot be adopted without critical analysis.

An object oriented approach generally provides good mechanisms for avoiding integration conflicts. However, when considering this approach, one must take into account that a single canonical model is required to describe the entire data model, which clearly restricts its flexibility and scalability. Each time a new stakeholder or data source enters the logistics system, the model needs to be extended. Depending on the dynamics of the logistics system, this may or may not be a disqualifying factor with regards to this approach. As the fluctuation of data sources, stake holders and systems in a complex logistics system with any degree of autonomous control can be assumed to be high, an object-oriented approach to data integration is likely to be unsuitable.

A loosely coupled approach requires detailed knowledge of each of the heterogeneous data sources to be able to be successfully employed. With regards to complex logistics systems, further analysis is required to determine whether this is feasible or not. The possibility of requiring highly flexible, and thus possibly not always pre-determinable, context data, for example from sensor networks, may prove to be an argument against this approach.

Besides the aforementioned traditional approaches to data integration, a number of predominantly semantic approaches remain to be taken into account. Here, the main concepts constituting architecture of such data integration systems are mediators (Ullmann 1997). In this approach, both syntactic and semantic descriptions of the data to be integrated are applied. The semantic mediator is capable of extracting knowledge regarding the data structures of the underlying data sources and subsequently transforming, decomposing and recomposing data requests according to that knowledge. The mediator relies on semantic descriptions of the data sources. In the case of item-specific product information management, this implies a wholly semantic modelling of product information and data across the distributed, heterogeneous sources, for which a number of approaches, such as ontologies, may be chosen. Here, extensive research is required to determine whether such semantic product models are feasible and adequate to address the requirements of item-specific product data management. However, the application of the semantic mediator to the problem area currently offers the most promising solution candidate — the following paragraphs attempt to sketch a technical solution to the problem on that basis.

Figure 3 introduces a possible solution architecture combining mediator technology with a service-oriented design pattern. The goal of this architecture is to provide a standard interface towards freely definable logical views of the data describing logistics entities which may be stored in arbitrary data repositories, including mobile, dynamic sources such as sensor networks, embedded systems etc.

The core of the suggestion is the provision of transparent mechanisms for the transformation between different standards and data structures. This is represented in Figure 3 below in the upper section by “data transformation and integration”, which can technically be realised by means of a suitable implementation of the mediator concept. To the left and right of that component, the semantic descriptions of the data sources are shown along with the necessary transformation rules by means of which the mediator is able to understand and decompose queries, forward them to the data sources in question and recompose a logical view of the returned data. Obviously, the nature of complex logistics systems demands a close analysis of system requirements with regards to the synchronisation of data access and maintenance of data persistence across the highly varied and dynamic data sources involved.

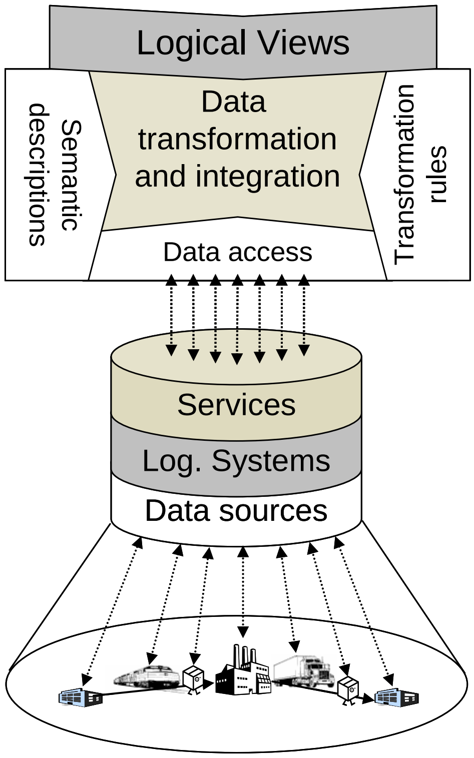

Figure 3: Suggested Approach to Data Integration

Access to both the logical views by the querying entities as well as to the individual data sources by the mediator is proposed by means of a service-oriented architecture. An architecture is described as service-oriented (SOA) when it uses loosely coupled software services to provide functionality (Stojanovic et al, 2004). Here, logic is not packaged as individual programmes, but is distributed across an amount of independent services. The actual implementation details of these services by their provider are completely transparent to the consumer. Currently, the most popular implementations of SOA are carried out using Web Services (Thoben et al 2003) which are built using the combination of the XML standards SOAP (Simple Object Access Protocol, cf. Gudgin et al, 2003), WSDL (Web Service Definition Language, cf. Christensen et al, 2001) und UDDI (Universal Description, Discovery and Integration, cf. Clement et al, 2004). These standards, in combination with the Hypertext Transfer Protocol (http), provide a system independent approach to the discovery, identification, provision and consumption of services according to SOA. However, a SOA may also be built using other technology such as CORBA (Common Object Request Broker Architecture), DCOM (Distributed Component Object Model) or Enterprise Java Beans (Blanke et al, 2004).

Approach and Methodology

Figure 4 below shows a matrix of methods and processes, the objectives to be reached by employing them, the standards and norms that can be utilised to support the methods and concrete tools that can be used. Three methods and processes have been identified which represent major areas of work towards the realisation of the suggested approach to data integration – the evolutionary development of the overall system, the application of ontologies for the semantic descriptions of data sources, and finally, methods and processes of software engineering to drive forward the technical implementation of the system.

Methods and processes	Evolutionary system development	Ontologies	Software Engineering
Objectives	Iterative approach	Service modelling, Description of semantics	Requirements definition, system conception, architecture
Standards and Norms employed	RUP	OWL, RDF, DAML+OIL, WSMO, usw.	UML 2.0
Supporting tools	MS Project	Protege	Poseidon

Figure 4: Matrix of Methods, Processes, Standards and Tools to be Employed

The overall approach to the development of the proposed solution should be one of an iterative nature. Several standardised development processes support this type of approach, foremost among these being the IBM Rational Unified Process (RUP). RUP is not a single concrete prescriptive process, but rather an adaptable process framework, from which the appropriate elements of the process can be selected. Several de-facto standard software tools exist to support project management in this context, one example of which is Microsoft Project.

Technologies and methods for semantic descriptions play a major role in the development of the proposed data integration approach. On the one hand, they can be applied to service-oriented architectures, with the aim of modelling them and the individual services from a semantic perspective. According to (de Bruijn 2005), Web Services, for example, can be modelled by the Web Ontology Language (OWL) based Web Service Modelling Ontology (WSMO) in combination with the Web Service Modelling Language (WSML).

The second major area in which ontologies can be applied in the context of this work is the semantic description of the individual data sources as well as to describe transformation rules. A number of formal languages are available for the definition of ontologies adequate for these purposes. The majority of these are based on mark-up languages such as XML. DAML+OIL (Darpa Agent Markup Language + Ontology Inference Layer, cf. Harmelen et al 2001) is an agent mark-up language for the semantic description of ontologies, which is based on XML and RDF (Resource Description Framework). The further development of DAML+OIL was taken up by the Web Ontology Language (OWL), which is the specification of a formal language for the description of terms and domains. OWL was developed to be machine-readable, that is, that software, such as agents, is able to understand the semantic descriptions formalised in OWL (Smith et al 2004). These properties make OWL a logical candidate for use in the development of the semantic descriptions of data sources in the proposed approach to data integration for complex logistics systems.

Conclusion

Currently today's logistics systems neither exploit their full potential nor do they fulfil the requirements of tomorrows dynamic logistics systems. Although powerful technologies and ICT-systems are available only local optima regarding efficiency, robustness etc. can be reached. The lack of relevant logistics data and information must be seen as the main reason for this situation. Business decisions are taken based on more or less local data and information. This lack of relevant data and information can be overcome by an approach for seamless data integration. It grants access to relevant data and information covering the whole value chain and therefore supports an intelligent decision making.

While this would mean a significant improvement of today's logistics systems it has to be considered as an absolute must for future organisational approaches like autonomously cooperating logistics environments.

Due to the diversity of existing ICT-solutions in the field of logistics a suitable data integration approach requires a unique access to different data sources while furthermore dealing with various still changing and evolving data formats and standards. Thus there is a need for a common interface which is generic enough as well as for structural knowledge supporting the identification and interpretation of data within the variety of different sources. The combination of a service oriented interface for data access with an intelligent mediator-based approach for the interpretation and transformation of logistics data as it is proposed here offers the flexibility which is required in the context of logistics systems.

References

Hannus, M.: Islands of automation in construction. In Žiga Turk, editor, Construction on the information highway, number 198 in CIB publication, University of Ljubljana, 1996.

Ballnus, R.: Erfolg mit EDI und E-commerce - Handlungsempfehlungen für die Abstimmung und Organisation interorganisationaler Netzwerke. Tectum Verlag, Deutschland, 2000.

Blanke, K., Krafzig, D., Slama, D.: Enterprise Soa: Service Oriented Architecture Best Practices, Prentice Hall International, 2004.

de Bruijn, J. et al: Web Service Modeling Ontology (WSMO), W3C Member Submission, World Wide Web Consortium, 2005.

Christensen, El., Curbera, F. Meredith, G., Weerawarana, S.: Web Services Description Language (WSDL) 1.1. W3C Recommendation, World Wide Web Consortium, http://www.w3.org/TR/wsdl.html, 2001.

Clement, L., Hatley, A., von Riegen, C., Rogers, T. (Hrsg.): UDDI Version 3.0.2, http://uddi.org/pubs/uddi_v3.htm, OASIS Open, 2004.

Gudgin, M., Hadley, M., Mendelsohn, N., Moreau, J.-J., Nielsen, H. F.: SOAP Version 1.2 Part 1: Messaging Framework, W3C Recommendation, World Wide Web Consortium, http://www.w3.org/TR/soap12-part1/, 2003.

Harmelen, F., Patel-Schneider, P. F., Horrocks, I. (Hrsg.): Reference Description of the DAML+OIL Ontology Markup Language. Joint United States / European Union ad hoc Agent Markup Language Committee. 2001.

Smith, M. K., Welty, C., McGuinness , D. L. (Hrsg.): OWL Web Ontology Language Guide, World Wide Web Consortium, 2004.

Stojanovic, Z., A. N. W. Dahanayake, H. G. Sol: Modeling and Design of Service-Oriented Architecture. In Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics: Impacts of Emergence Cybernetics and Human-Machine Systems. IEEE Computer Society, 2004.

Ullman, J.D.:.Information integration using logical views. In F.N. Afrati and P. Kolaitis, editors, Proceedings of the 6th International Conference on Database Theory (ICDT'97), 1997.

Wache, H.: Semantische Mediation für heterogene Informationsquellen. Akademische Verlagsgesellschaft, Dissertationen zur Künstlichen Intelligenz, Berlin, 2003.

Yergeau, F. Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E.: Extensible Markup Language (XML) 1.0 (Third Edition) - W3C Recommendation. World Wide Web Consortium, (Massachusetts Institute of Technology, European Research Consortium for Informatics and Mathematics, Keio University), 2004.