Improving Business Information Exchange using Semantics

Albert BOKMA[1], Stuart CAMPBELL[2], Sabina EL HAOUM[3], Kevin GINTY[4], Matt HINDS[5]
[1]*University of Sunderland, St Peter's Campus, Sunderland, Tyne and Wear, SR6 0DD, UK*
Tel: +44 191 515 3233, Fax: +44 191 515 3236, E-mail: albert.bokma@sunderland.ac.uk
[2]*TIE, Beech Avenue 180, 1119 PS AMSTERDAM (Schiphol-Rijk), The Netherlands*
Tel: +44 1270 254 019, E-mail: stuart.campbell@tieglobal.com
[3]*Carl von Ossietzky University of Oldenburg, 26129 Oldenburg*
Tel: +49 441 798 44 76, Fax: +49 441 798 44 72, E-mail: elhaoum@wi-ol.de
[4]*University of Sunderland, St Peter's Campus, Sunderland, Tyne and Wear, SR6 0DD, UK*
Tel: +44 191 515 3270, Fax: +44 191 515 3236, E-mail: kevin.ginty@sunderland.ac.uk
[5]*University of Sunderland, St Peter's Campus, Sunderland, Tyne and Wear, SR6 0DD, UK*
Tel: +44 191 515 2804, Fax: +44 191 515 3236, E-mail: matt.hinds@sunderland.ac.uk

Abstract: E-Business information systems require effective inclusion of semantics to function. Emerging messaging systems, such as ebXML, need to include structural support for these semantics. In order to develop an ontology based approach to this task, we have undertaken several case studies and used the results to derive generic use cases for the ontology based e-business semantics development. This paper describes the results of these case studies and the comparison of sectors that will be used to inform the generic use case development. Finally we propose a solution to this semantic problem that will enable the publication and synchronisation of business messaging semantics as utilised by participating organisations in a distributed web architecture.

1. Introduction

In recent years, the ultimate strategy for companies seems to have been the concentration on the things they know best – core business. At a macroeconomic level, with slow revenue growth, the management of partner related expenditures provides an opportunity to boost profits through cost control. One example is the outsourcing of all non-core competencies to specialised companies with particular expertise promising better quality services at lower costs. As steps such as these escalate in importance, the significance of managing electronic relationships with these partners increases dramatically. To improve or sustain financial performance, supply chains are forced to organise themselves in networked electronic structures. The key success factor for these networked businesses is the efficiency of their operations so that the potentials of the partnership are not swept away through complexities, difficulties and the overall costs of implementation.

To enable these relationships, key themes that must be addressed are identified in Figure 1. From this diagram a number of key themes are identified and semantics is one of the key building blocks to this framework structure. What is more important however, is that it underpins all other layers. Only communication, which is now technical, ubiquitous and in fact a commodity item, crosses all themes. The remainder of the stack can be considered as business items, and at every other level (content, process, repository) they are built from semantic constructs. Their use and interoperability will be a key factor in determining the cost, benefits and return on investment (ROI) of any system.

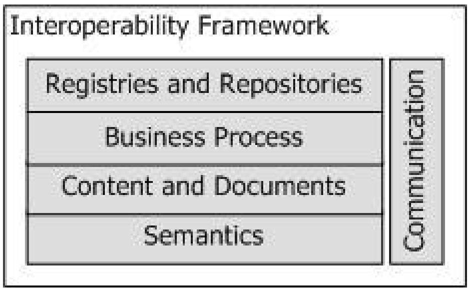

Figure 1: Role of semantics in an interoperability framework [21]

Business is now recognising that it needs efficient and global infrastructures to share semantics between partners and also to use them within internal interoperability situations – even internal applications talk in different dialects! Technical solutions are needed which make semantics truly part of an eBusiness architecture.

Standards like Electronic Data Interchange (EDI) are used to implement these solutions and are well defined with respect to protocol and semantics. However, dynamic business interactions require variable and adaptable protocols such as webservices, which EDI is unable to provide. The specifications used for these protocols cover the methods of communication but not the what, semantics is missing. To fill this gap, the ebXML (electronic business extensible markup language) standard can be used as it already provides semantic definitions by providing standard types (the Core Components) for type based semantic composition.

2. Objectives

This paper is based on work that is currently part of a jointly funded research project (partially funded by the European Community 6th Framework programme under contract FP6-2005-IST-5-034980). The overall objective of the paper is to propose the structure for a new application that will enable the use of semantics within supply chain integration. To achieve this, we present results of case studies that have been performed with companies from two diverse sectors, furniture and automotive, across two continents (Europe and Asia). The case studies were designed to help consider the realities of how a semantically enabled solution might be implemented.

3. Case Study Structure

As mentioned previously, case studies were performed across two diverse sectors. To ensure the case studies were productive it was decided early on that a focused approach would yield the best results. To facilitate this, a questionnaire was created with a blocks of questions targeted at specific groups. These groups were Management; Ordering and Purchasing; Supplier/Provider and Customer/Buyer relationships; Security and legal Issues; IT Issues [1]. This grouping enabled the interviewer to ask questions relevant to the person they were interviewing. Depending on the response the interviewee gave, the interviewer was then able to ask follow on questions to obtain as much information as possible. All interviews were performed in a face-to-face environment.

It was recognised that concepts such as ontologies and semantics may not be understood in the majority of companies. To overcome this potential problem, interviewees were given a short presentation to describe these concepts in simple terms. Although the results are discussed in more detail in the next section, the general response was a general understanding was reached, allowing a more detailed and productive interview process.

Two companies from each sector were identified for interview. It was also decided that within each sector a company from the EU and another from China would be selected. This choice allowed us to build up a picture of how companies across continents communicated.

Within the automotive sector ZF Friedrichshafen AG from Germany and Beiqi Foton Motor Co. Ltd. (BFM) of China were selected. The selections from the furniture sector were Mariner from Spain, and Shanghai Sunline from China. During the interviews, further organisations that supplied or provided a service to these companies were identified and interviewed. This enabled us to build an accurate picture of how the inter-company relationships operated.

4. Case Study Analysis

The interview results from the Automotive sector showed that it was based on the model illustrated in Figure 2. At the top of the model are the Original Equipment Manufacturers (OEM), these companies rely on the tier 1 suppliers who provide the components or modules. In turn these suppliers rely on lower tier suppliers for some products. To enable these supply chain communication channels the sector, within the EU, relies on EDI heavily, and have done so for many years.

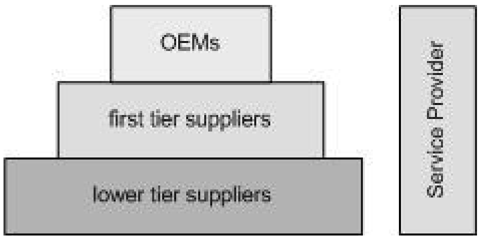

Figure 2: Traditional span of services [1]

A problem encountered within these supply chains is that there are a wide variety of document standards used. This interoperability problem is caused by the diversity of the IT-systems present and is a major problem for SMEs. These organisations are generally located within the lower tier suppliers of Figure 2 and as such few have direct links back to OEMs. They also do not have the capacity to employ a dedicated IT resource who can handle these problems as the costs can be quite prohibitive. To overcome these issues a service provider can be used to handle the communication between tiers.

Within China a slightly different communication model is encountered. EDI has not yet made the same impact as it has across Europe. Instead the large suppliers open up dialogue at the contract stage to discuss terminology. With BFM this generally results in the supplier using the same terminology.

One important point to note is that the automotive sector interviewees currently have no cross continent business. However, this is an area that they do want to move into in the future.

The furniture sector is a complete contrast of the automotive sector. Whilst there is a well defined supply chain within the automotive sector, very few exist within the furniture sector. Instead direct relationships between companies are common. The industry is also predominantly based around SMEs, resulting in substantial diversity where IT systems are concerned. Until recently the industry has been driven by human relationships and communication has been performed using traditional methodologies. The industry is only now starting to embrace new technology, for example, customers are requesting electronic catalogues forcing the supplier to cut the time lag for providing product information.

Communication between key actors within the industry is disjointed. Systems such as ERP (Enterprise Resource Planning), are not common within the industry (for example only 23% of SMEs in Spain have implemented one) [2]. This leads to key documents being passed through different channels, including E-Mail, Fax and regular mail. The result of this is, and is not limited to, lost orders/documents and problems during production [1].

Within the furniture sector competition from Asia has increased, this has lead to some manufacturers extending activities such as manufacturing to other countries. Mariner has done this by going into partnership with Sunline, both of which were interviewed for this research. This relationship has identified issues beyond basic interoperability that any semantically enabled solution will need to address. An example is shown in Table 1.

Table 1: Simple translation of the Pinyin word mu to Spanish and English

Reference	Pinyin	Spanish	English
A	. mǔ (s)	1. madre. 2. hembra	1. mother. 2. female
B	. mú (adj,)	Molde	Mould
C	. mù (s)	1. árbol. 2. madera. 3. ataúd. 4. entorpecido	1. tree. 2. wood. 3. coffin. 4. obstructed
D	. mù (s)	1. ojo. 2. ítem / punto / artículo. 3. orden. 4. lista / catálogo	1. eye. 2. item / point / article. 3. order. 4. list / catalogue

Table 1 shows the differences that are encountered trying to find the meaning of the pinyin word mu. Whilst this is a simple example, it illustrates the problem that semantic reconciliation will face. These problems are not restricted to cross continent communications, they are also present in communications within Europe, as illustrated in Table 2. This table illustrates the results of a literal translation from Spanish to English and what is used normally for the English description. Again a semantically enabled solution will need to overcome this type of problem to be successful.

Table 2: Translating from Spanish to English

Spanish Description	Textual translation in English	English Description
Cara frontal	Frontal Face	Front Panel
Proforma/nota de pago	Quote/ payment confirmation	Quote/"proforma" invoice

5. Use Case Definition

Once the case studies were completed the next step to deriving a potential solution was to define specific use cases. These case studies focused on the communication between actors and the methods currently used. These case studies are defined in [2].

Once analysis of these case studies was complete it was clear that although the sectors were diverse, some of the problems existed across both sectors. For example, taking the example from the automotive sector regarding EDI Communications to Tier 1 Suppliers. The same scenario existed in the furniture sector, albeit not using EDI as the communication method. The generic use case for this scenario is therefore that organisations need a tool that allows this type of message passing, using a commonly understood standard. Full results for the generic use cases can be found in [2].

From these results four key areas that a solution required were derived. Firstly, a semantic specification will need to be defined, this can be created from existing documents that a user has. Secondly, the specification will need publicly published to allow other companies to see this specification. In addition to publishing a facility to retrieve will also be required. Thirdly, to allow automatic matching and cross referencing a negotiation and pooling facility is required. This will allow companies to automatically link their semantic specification to another companies and allow business relationships to be formed quicker than is presently possible. Finally, as the system expands the quality of the information may degrade as wrong links could be made between specifications. A semi-automated system is required that will allow identification of these inconsistencies and notify users who can then remove them if required.

6. Implementing a semantically enabled solution

The four key areas identified at the end of the previous section are listed in Table 3. Against each area are specific software elements that need to exist in our potential solution. The integration of these elements are depicted in Figure 3 and described in more detail thereafter.

Table 3: Deriving a possible solution from the case studies

Area	**Software element**
Semantic Specification	A Semantic Analysis Tool (SAT) which will enable a user organisation to feed in their current semantic/structures and analyse them
Publishing and Retrieval	These utilities can be used to create individual Beacons (using registries and repositories) The Viewer will allow any party to browse the network to examine particular semantic structures.
Negotiation and Pooling	The beacons will operate in a federated manner as the Flock. A Comparator will be provided which can be utilised by the researching organisation to determine semantic possibilities (using the SAT and Flock basics).
Quality Assessment and Analysis	Conflicting information will occur from time to time as the incorrect mappings will have been made (either automatically or manually), the Distributed Information Purifier supports cleansing and consistency.

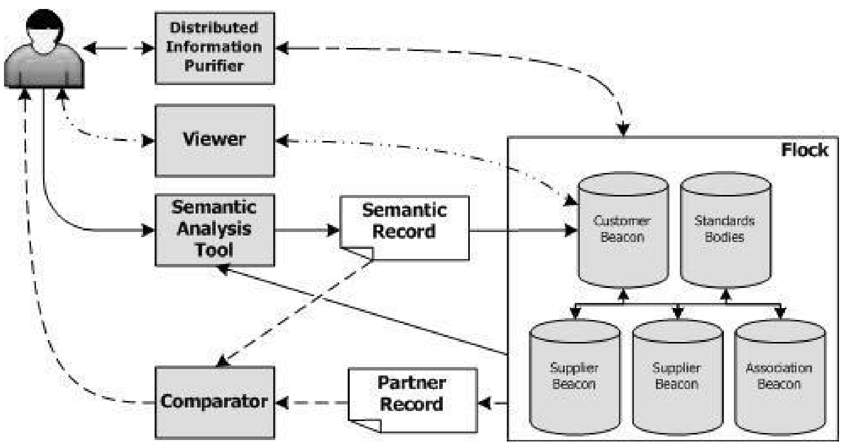

Figure 3: Semantic Solution Concepts

- At the Core of the system are components called Beacons. These act as a local store for semantically tagged information and are created using other tools. They are deployable in private organisations, sectoral associations or international community organisations which similarly semantically tag information such as core specifications, implementation guides and classification trees.

- The beacons will operate in a federated manner called the Flock. In subsequent stages unknown or uncertain semantic relationships can use the knowledge in all Beacons to help build semantic pathways. This feature will utilise distributed Registry and Repository functionality which is already present in the EU SEEMseed project [15].
- The Semantic Analysis Tool (SAT) will enable a user organisation to feed in their current semantic/structures (e.g. XML, EDI, Flatfile, RDB based structures). This tool will analyse and semantically tag these based on user (e.g. internal to an organisation), sector (e.g. pointers to sectoral de facto standards like RosettaNet) and global semantic knowledge (e.g. ISO and UN code lists) as well as knowledge in the flock. Finally this information will be placed into the flock via the customer beacon.
- To support organisations needing to access each other and determine their semantic compatibilities a Comparator will be provided which can be utilised by the researching organisation to determine semantic possibilities (using the SAT and Flock basics). In case of unclear results, the user is prompted in order to decide manually, allowing both high automation and high quality. When resolved these relationships will be updated back to relevant beacons. This will enable content based semantic interoperability.
- The Viewer will allow any party to browse the network to examine particular semantic structures and their use. This is particularly valuable when establishing new content flows such as messages or interface definition. Allowing the existing strategies and definitions to be taken advantage of. This will in turn bring the user benefits since these semantics are already part of the network, facilitating the establishment of a relationship with a second company.
- It is feasible that there will be conflicting information within the Flock, as sometimes the wrong mappings will be made. This will result in the data and relationships needing to be cleansed. The Distributed Information Purifier will perform this task in an automated process, with human intervention when required.

This solution proposes a semantic grouping and organization of content approach in P2P networks with several novel aspects. Firstly, a semantic peer is not limited to be a simple data source but it is a complex mediator system, which integrates heterogeneous data coming from different data sources. Moreover, in contrast to the recently announced semantic overlay network approach [3] designed for flat, pure peer-to-peer topologies and for limited meta data sets, such as simple filenames, we allow a grouping of complex heterogeneous data sources, including textual and multimedia documents. Finally, another aspect to be investigated is the case when a data source of a semantic peer is another semantic peer, introducing the concept of a semantic super-peer.

With reference to mappings, it is worth stating that the mappings among the different concepts are an uncertain and vague concept, not easily associated to an absolute value. There is a certain degree of similarity between two concepts, which varies from 0 (completely different) to 1 (perfectly equal) and the several degrees of similarity are not perfectly bound. The degree of membership of the "similar" concept can vary depending on the observer and context. Uncertainty becomes prevalent in the network, where more local ontologies are involved. It is often the case that a concept defined in one ontology can only find partial matches to one or more concepts in another. To overcome the difficulties arising from the crisp logic, existing ontology languages (e.g. RDF [4], OWL [13]) need to be extended to be able to capture uncertainty knowledge about the concepts, properties and instances in the domain and to support reasoning with partial, imprecise information. With our solution, we propose to tackle this problem by adopting an approach which extends both the ontology definition language and the knowledge processing, such as schema matching and query propagation and execution. Existing schema matching algorithms as the one proposed in [10]and [11] could be adapted for this context.

Multi-lingual support within this solution will build on work taken from CLEF and elsewhere including document translation (see [8]) and statistical query-translation (described in [19], [20]). This will support concept translation and the development of query reformulation techniques to query peers in different languages.

Methods and tools for managing the dynamics of the virtual network will also be provided. To this end, principles of data schema evolution (described in [9], [17]) and versioning (see [6], [14]) are reusable for ontology evolution and versioning respectively (see [18]). Among the few works useful for this solutions purpose in this rather new field of ontology evolution are [16], which provides different levels of granularity for change operators, resulting in a taxonomy of compound change operators; [7], which focuses on ontology evolution on the Semantic Web; [5], and [12]. Some approaches are thus available, but they are incomplete and lack a general approach. None of these methods have been tested thoroughly in real case studies.

7. Conclusions

In conclusion, this paper has discussed the structure and results of the case studies we conducted to help us to create a possible semantic solution to business interoperability. From these we feel it is clear that there are numerous advantages to integrating semantics into the supply chain. The case studies also identified that SMEs are not interested the theories behind these applications, i.e. ontologies or semantics. They are more interested in expanding their business and making interoperability easier.

The information obtained from these case studies allowed us to propose a solution that attempts to hide the complexity of semantics and ontologies from the end user. Whilst developing these solutions it will be important to keep in mind that users do not want to be exposed to the intricacies involved. Ultimately the tools will enable the companies involved, as well as other companies, to trade electronically with other partners, including the larger corporations.

Work is currently being carried out on the design and implementation of the solution defined above. As only the initial stages of this process have been completed, we are unable to verify the solution against the case studies to ensure that it works. Further work with the interviewees will be carried out once the software is at a stage to be tested. This will help us to identify any shortcomings or problems and will be the topic of future papers.

References

[1] Bhullar G, Núñez MJ, Roca de Togeras A, Garcia O, Campbell S, (2007) STASIS D2.2.1 Target Market Sector Report [online]. Available from: http://www.stasis-project.net/deliverables/ [25th April 2007].

[2] Núñez MJ, Garcia O, Delgado M, Roca de Togores A, Bhullar G, Campbell S, (2007) STASIS D2.2.2 Use Case Report [online]. Available from: http://www.stasis-project.net/deliverables/ [25th April 2007].

[3] Garcia-Molina H, Crespo A, (2003) Semantic Overlay Networks for P2P Systems. Technical report, Computer Science Department, Stanford University.

[4] Haase P, Broekstra J, Eberhart A, Volz R, (2004) A Comparison of RDF Query Languages. In Proceedings of the Third International Semantic Web Conference, 7-11 September 2004 Hiroshima. Berlin: Springer, 502-517.

[5] Heflin J, (2001) Towards the Semantic Web: Knowledge Representation in a Dynamic, Distributed Environment. Thesis (PhD). University of Maryland, Collega Park, MD, USA.

[6] Kim W, Chou H, (1988) Versions of Schema for Object-oriented Databases. In Proceedings of the 14th Int'l Conference on Very Large Data Bases (VLDB88) 29 August - 1 September 1988 Los Angeles. Morgan Kaufmann, 148-159.

[7] Klein M, Fensel D, Kiryakov A, Ognyanov D, (2002) Ontology Versioning and Change Detection on the Web. In Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02), 1-4 October 2002 Siguenza. Springer-Verlag, 197-212.

[8] Lam-Adesina AM, Jones GJF, (2003) Exeter at CLEF 2003: Experiments with Machine Translation for Monolingual, Bilingual and Multilingual Retrieval. CLEF Workshop.

[9] Lerner B, (2000) A model for compound type changes encountered in schema evolution. ACM Transactions on Database Systems (TODS), 25 (1), 83-127.

[10] Mandreoli F, Martoglia R, Tiberio P, (2004) Approximate Query Answering for a Heterogeneous XML Document Base. In Proceedings of the 5th WISE Conference.

[11] Nottelmann H, Straccia U, (2005) sPLMap: A probabilistic approach to schema matching. 27th European Conference on Information Retrieval Research. ECIR 2005.

[12] Oliver D, Shahar Y, Musen M, Shortliffe E, (1999) Representation of Change in Controlled Medical Terminologies. AI in Medicine. 15 (1), pp.53-76.

[13] OWL Web Ontology Language Overview, World Wide Web Consortium W3C 2004

[14] Roddick J, (1995) A Survey of Schema Versioning Issues for Database Systems. Information and Software Technology. 37 (7). pp.383-393.

[15] SEEMSeed Project. http://www.seemseed.net [12th October 2006]

[16] Stojanovic L, Maedche A, Motik B, Stojanovic N, (2002) User-driven Ontology Evolution Management. In Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02), 1-4 October 2002 Siguenza. Springer-Verlag, 285-300.

[17] Banerjee J, Kim W, (1987) Semantics and Implementation of Schema Evolution in Object-oriented Databases. In Proceedings of the 1987 ACM SIGMOD international conference on Management of data, 27-29 May 1987 San Francisco. New York: ACM Press, 311-322.

[18] Bergamaschi S, Castano S, Beneventano D, Vincini M, (2001) Semantic Integration of Heterogeneous Information Sources. Special Issue on Intelligent Information Integration, Data & Knowledge Engineering. 36 (1). pp.215-249.

[19] Federico M, Bertoldi N, (2002) Statistical cross-language information retrieval using n-best query translations. In Proceedings of 25th ACM SIGIR conference on Research and Development in information retrieval, 11-15 August 2002 Finland. New York: ACM Press, 167-174.

[20] Gao J, Nie JY, Xun E, Zhang J, Zhou M, Huang C, (2001) Improving query translation for CLIR using statistical models. In Proceedings of 24th ACM SIGIR conference on Research and Development in information retrieval, 9-14 September 2001 New Orleans. New York: ACM Press, 96-104.

[21] Campbell et al., (2005) Original STASIS Proposal