MOTIVATING ENGINEERING/BUSINESS PROBLEM

The AEC supply chain is fragmented in the same way as most information management processes in the industry are. An integrated supply chain would include activities from design, procurement to installation, performed by different parties and organizations. As the current practice indicates, having information available as needed can significantly reduce lead-time as well as increase accountability for tracking purpose. Sharing of information within and across companies is critical in effective supply chain management [Lee and Whang 2000]. However, this is not an easy task: the information sources are often scattered in several locations, utilizing different software and hardware platforms and not easily accessible. Information gathering takes too long; no single individual can handle all of information in a supply chain because one has to deal with many products, numerous requirements, and many types of transactions. Another difficulty in sharing information is the nature of the AEC business: "One-of-a-kind nature of project, temporary multi-organization". Specifying proprietary-designed representations and "one-time" information channel to exchange data and knowledge is not viable or justifiable. In addition, different goals and objectives of project participants deter information sharing. If these difficulties can be overcome, there is a lot to be gained in the AEC industry.

The purpose of interoperation is to increase the value of information when information from multiple and, likely, heterogeneous sources can be accessed, transferred and integrated. Interoperability allows two or more information sources to exchange information and to re-use the information for further purposes. Interoperation therefore adds values to each individual source and enhances efficiency and productivity in a supply chain. The goal is to reduce the barriers between organizations and to open for collaborative supply chain. A study performed by NIST reports that inefficient interoperability costs more than $15.8 billion, in the year 2002 alone, to the construction industry on the design, construction and maintenance of large commercial, institutional and industrial buildings (not including public works and other civil infrastructures systems) [Gallaher et. al. 2004]. An earlier study by NIST Strategic Planning and Economic Assessment Office conservatively estimates the economic cost due to lack of interoperability in the United States automotive supply chain alone at one billion dollars per year [NIST 1999]. The trend of creating virtual supply chains due to the advent of computers and Internet signifies the importance of information interoperability in various industries including the engineering and construction industry.

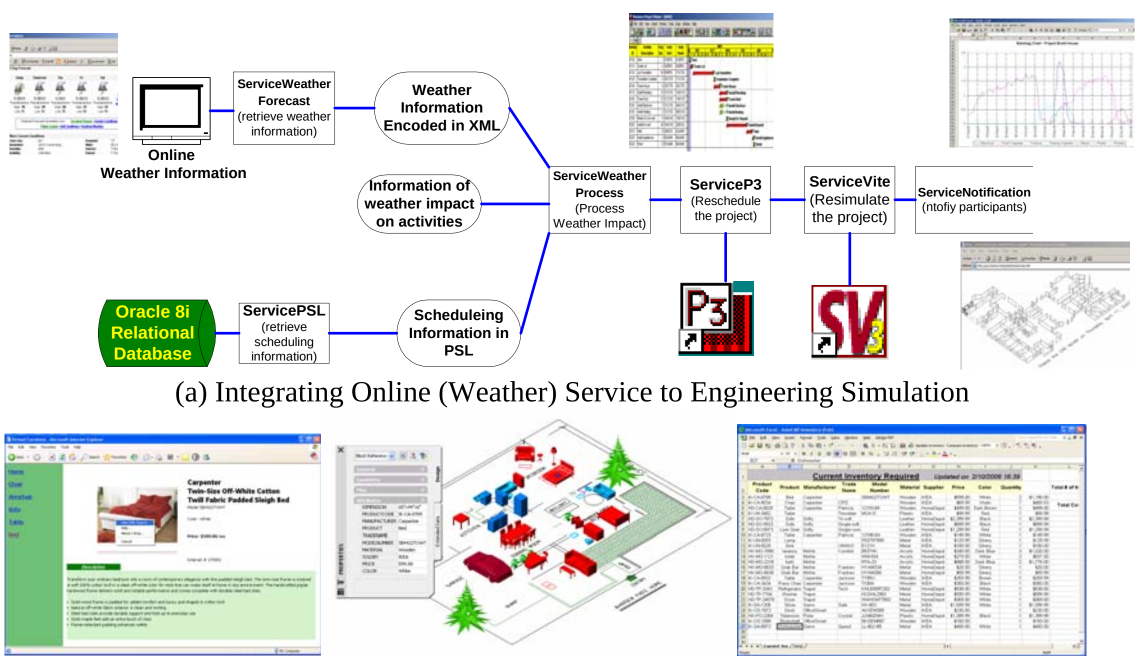
This seed research proposal is to continue the study of information interoperability for AEC supply chain applications to connect and to integrate distributed information sources. The purpose is to create a framework that can coordinate distributed, semantic-rich engineering services in order to solve the fundamental interoperability problems faced by the AEC industry and to support workflow and supply chain applications. In this coming year, our focus will be on the issue of information interoperation, and two deliverables will be produced: (1) mediation of heterogeneous information sources to discover similarity of concepts and to "harmonize" different ontological standards; and (2) formal operations for mapping and integrating information to support integrated services.

THEORETICAL AND PRACTICAL POINT OF DEPARTURE

The value of information sharing for supply chain management has been recognized in both research and practice (see, for examples, [Lee et. al. 2000, Coleman and Jun 2005]). Standards and approaches for information sharing such as Electronic Data Interchange (EDI) and RosettaNet have existed in manufacturing supply chain for over twenty years. However, composition of distributed systems and integration of proprietary information sources such as local databases that use heterogeneous data structures remain to be challenging tasks. The time and effort associated with setting up and maintaining these infrastructures pose a formidable impediment to their adoption, especially for small to medium-sized companies which most AEC companies are. Moreover, unlike many manufacturing supply chains that allow long-term relationship to be established between business partners, the relationship between stakeholders in AEC supply chain is often project-based and temporary. That is, AEC supply chains keep changing from project to project and companies cannot afford to spend much time in configuring a system. Therefore, an application for this industry has to be flexible enough to accommodate different supply chains efficiently. Effective and flexible information sharing and service integration is therefore required for supporting supply chain activities in the AEC industry.

On-Line Web Services

With the rapid development of Internet technologies, the computing environment is evolving toward an interconnected web of autonomous services, both inside and outside of enterprise boundaries. A "web service" can be described as a specific function that is delivered over the Internet to provide a service or information to users. Web service integration allows the automation of application-to-application or organization-to-organization cooperation using the Internet infrastructure. Fig. 1 shows two example scenarios to illustrate bringing on-line services to engineering simulations [Cheng et.al. 2006]. Fig. 1(a) shows the integration of online weather forecast information (represented in XML) with project information (expressed in PSL, an ISO standard for process language [Gruninger and Menzel 2003]). The impacts of the weather conditions to the schedules and resources can then be simulated using Primavera P3 and Vite SimVision. The results can be queried and displayed as charts using Microsoft Excel and web browser. As another example, Fig. 1(b) shows the possibility of dynamically downloading product information (from online catalog or suppliers' database) into a design and creating a procurement list (that can further be used for further product comparison and inventory management). In this example, Autodesk's i-drop technology is employed to drag-and-drop product description into a design. Manufacturing information such as model code, manufacturer, supplier, price quote, etc. are also attached as object attributes such that a procurement list (and cost estimation) can be produced during design. These two examples demonstrate, among many, the potential benefits of integrating web-based services.



(a) Integrating Online (Weather) Service to Engineering Simulation

(b) Integrating Online Catalog with CAD and Inventory

Fig. 1. Example Scenarios for Integration of Online Web Services

Many languages have been proposed to facilitate the development of Web services. Examples include Web Services Description Language (WSDL) [Booth and Liu 2004] and Flow Language (WSFL) [Leymann 2001], Business Process Execution Language for Web Services (BPEL4WS) [Andrews et al. 2003], and Web Service Ontology based on DARPA Agent Markup Language (DAML-S) [Ankolekar et al. 2001]. Recently, semantic web services modeling language such as WSML [de Bruijn et al. 2005] have been proposed to promote the use of semantic reasoning about services. This level of integration is the province of semantic web services and deals with semantic interoperation of application services [Preist 2004]. To integrate distributed services over the Web, information interoperability is critical, so that results can be reused by other applications.

Unified Information Models

Many efforts have been spent attempting to build a single, unifying standard model of concepts and definitions to capture various phases and facets of a supply chain. The earlier development of the IFC and the current use of Building Information Model (BIM) have been focused on establishing unifying models to describe product, process and organization information in aspects such as design, analysis, procurement and installation (even though individual applications would likely use certain aspect and only portion of the model). Such unifying models, however, are neither efficient nor practical [Ray 2002]. For example, while product information emphasizes on physical geometry, process and organization information focus on task precedence and organizational resources and structures. Information for analysis includes analytical requirements and results whereas information for procurement deals with materials supply and delivery. Differences in the scopes of information constrain the structure and the extent of information. Separate yet linked models differentiated by types and scopes are easier to manage and more flexible for information exchange between multi-disciplinary applications.

Some organizations have developed mechanisms to link and relate different data formats and models. For instance, NavisWorks currently tries to build the mappings among models for large number (>40) of applications. Although its endeavor appears to be worthwhile, such an approach, that attempts to establish direct one-to-one mappings between models, is simply not scalable as the number of applications and models increases. A scalable methodology to bridge the models through mapping is therefore necessary to facilitate the interoperation between separate models.

Information Mapping between Interoperability Standards

Even within a particular domain or industry, multiple terminological classifications or data model structures exist. For instance, in the AEC industry, there exist a number of information standards to describe the semantics of building models, such as the Industry Foundation Classes (IFC) [IAI 1997], the CIMsteel Integration Standards (CIS/2) [Crowley and Watson 2000], and the OmniClass Construction Classification System [CSI 2006]. Each of these standards is constructed for a specific purpose and from specific viewpoints. For model rebuilding and data exchange purposes, comparison and mapping between heterogeneous standards are often necessary. Although the standards share similar scopes of interest, their differences in terminology and viewpoints make the task of comparing and mapping difficult.

Mapping information from different sources is now done manually by domain experts. Our experience has shown that semantic mapping of two distinct ontology standards is a time consuming process [Cheng et.al. 2002]. A recent work by Begley et. al. [2005] on semantic mapping between ifcXML, representing the information content of IFC, and Automating Equipment Information Exchange (AEX) by FIATECH [Teague, Palmer and Jackson 2003] just for centrifugal pumps has taken considerable amount of efforts. Lipman [2006] spent about two years in an attempt to map between CIS/2 and IFC for structural steel, and concluded that some of the information in CIS/2 defining the analysis, design and detailed models is lost when mapping to IFC. These efforts show that manual mapping between standards is labor-intensive and time-consuming. This seed research attempts to tackle this fundamentally complex but important ontology mapping problem using a knowledge-driven mediation approach to harmonize the terms and attributes between heterogeneous ontology standards.

METHODS

Ontologies have been increasingly leveraged to describe the terminology and structures used in the information sources in order to facilitate the interoperation between them. It has been forecasted that "By 2010, ontologies ….will be the basis for 80 percent of application integration projects" [Jacobs and Linden 2002]. A ontology acts as a terminological basis for information exchange and provide a means for knowledge sharing and reuse. They define the concepts, the relationship between concepts, the concept instances and the axioms of the information to be exchanged, and serve as a tool to help understand the semantics and formats of information. In a web of ontologies, transforming information from one ontological standard into the context of the users and their applications is important [Wiederhold 1994]. The migration of data requires comparison of concepts, renaming of terms and even modifying the data structures. Tools are needed for the extraction of terms and "harmonizing" ontologies.

Formal operations are necessary for the articulation of overlapping ontologies (as depicted in Fig. 2).

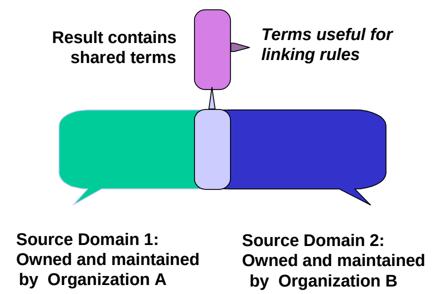


Fig. 2: Intersection operation to define articulation between two sources

Mediation has been shown to be a useful means to resolve problems of semantic differences among different information formats. Metadata are defined or extracted from each information source. To construct the needed metadata, advanced text and data mining methodologies can be deployed to analyze and compare the information sources, thus creating rules that can enable articulation, linking disjoint information resources, and interoperation. The mappings can then be used in applications such as query rewriting and information and knowledge correspondence. This seed research proposal intends to extend the mediation methodologies to reveal the semantic similarities and differences among the ontology standards and implement the mapping and transformation of heterogeneous ontologies. The research consists of two basic work packages:

- Develop knowledge-based approaches to compare the semantic similarities of ontology standards and to develop articulation rules for mapping
- Investigate methodologies to transform information about the same domain but from different ontology structures such that the information can be integrated across scopes and applications

Knowledge-Driven Approach for Semantic Similarity Comparison

Surveys on the various approaches for ontology mapping (merging, alignment) have been reported [de Bruijn et al 2004, Euzenat et al 2004]. The difficulty in comparing the semantics of terms and concepts is often underestimated. Different terminologies in two standards sharing the same meaning, for example, door hardware and door furniture, are usually observed. Identical terms yet with different meanings are also not uncommon due to the differences in domains and viewpoints. Manual recognition of semantically similar concepts performed by domain experts is regarded as the most trustworthy analysis. However, the process is subjective and multiple interpretations can result. Automated comparison based on the linguistic similarity between concepts is growing in popularity in recent years. Some common approaches include term matching that relates terms with the same words, synonyms, or terms with the same root. Dictionaries have also been used [Ehrig and Sure 2004, van Hage et. al. 2005] to help define and compare ontology concepts. However, the reliability is not guaranteed because the set of synonyms and the definitions could be different from different sources. In addition, many concepts have different meanings when used in different contexts. For example, the concept "*finishes*" often refers to the decorative texture or coating of a surface in a construction project whereas it means to complete or to terminate in a general sense.

We take a different approach to compare the concepts of heterogeneous ontologies by using a knowledge base in a domain similar to the domain of the mapping ontologies as a bridge. Fig. 3 illustrates an example of matching the concepts from ifcXML and those from OmniClass, which incorporates the contents from standards such as MasterFormat and UniFormat. Both groups of concepts are latched to a set of textual documents in International Building Code (IBC) [ICBO, 2006] which acts as a knowledge base. Advanced text mining and statistical analysis techniques are deployed to discover the knowledge of semantic similarity among concepts. With the intuition that related terms should appear in the same paragraphs or sections (using the building code as a corpus), concept comparison and matching by co-occurrence is proposed to map different sets of terms from heterogeneous ontologies. The co-occurrence frequency of two concepts in the set of documents reveals the closeness of the two topics and acts as a means to compute the relatedness between them. Our experience so far has revealed that the set of documents utilized should be in the same (overlapping) domain as the mapping ontologies in order to capture the semantically related and overlapping concepts.

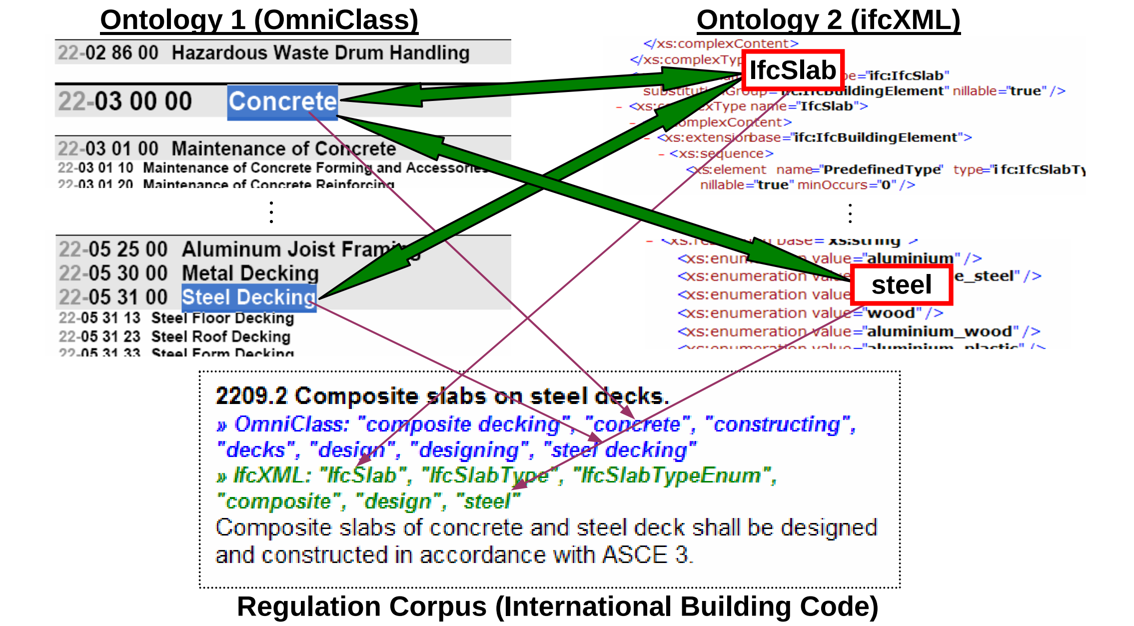


Fig. 3: Mapping between OmniClass and ifcXML using IBC as the Knowledge Base

Information Transformation and Integration

Once semantic similarity among the original and targeted ontologies has been identified, information transformation can be performed on models from one ontology to another. It includes the migration of data between standards in similar domains such as MasterFormat and UniFormat, and between proprietary standards and open standards. Both may involve one ontology significantly richer than the other, leading to potential loss in information during transformation. Integration of information across domains is unavoidable in the multi-disciplinary AEC supply chain. For example, the connection between green building information and product geometry information is gaining attention in the design and construction industry. The overlapping portions of the ontologies from two different sources need to be identified and a mapping between them built before integration becomes feasible. All these transformation and integration can be facilitated by defining a mapping language, which works as a mediator to map between heterogeneous standards.

Identifying the semantic similarity is necessary but not sufficient for information mapping. The "Form, Fit and Function" evaluation approach suggested by ANSI [2002] can be used to obtain more accurate mapping between different information formats. Form refers to the physical structure and content, fit covers the item's semantics or meaning, and function concerns the purpose or how the entity is used. When all three conditions are equivalent, the entities are considered as equal; for instance, a company is called a seller in one application but a supplier in another. One common situation in materials procurement is the ambiguous interpretation of terms such as "product cost". Product cost in the buyer's perspective means the money spent on obtaining the product from suppliers and storing it. On the contrary, product cost in the supplier's perspective can represent the money spent on producing and keeping stock of the product. In this case, the two interpretations have the same physical form and semantic meaning, but not the business function. At the same time, product cost in the supplier's point of view can also represent the price it sells the product, which share the same physical form and business function with the buyer's interpretation but not the semantic fit. Therefore, evaluation based the form, fit and function metrics of data can potentially provide further validation on the correct matching of information.

RELATIONSHIP TO CIFE GOALS

This research focuses on information interoperability and services integration across multiple distinct applications in a construction supply chain. This proposed research touches upon at least five of the CIFE research areas and goals:

Engineering modeling methods:

- Modeling of product, process and organization data for information exchange and sharing purpose.

Analysis methods:

- Establishment of formal methodologies and operations to discover and to compare semantically equivalent or related concepts between different ontology schemas and standards.

Business metrics:

- Support for information flow and supply chain applications.

Strategic management:

- Integration of loosely coupled and distributed information and services for strategic decision-making.

Economic Impact analysis:

- Cost saving due to efficient collaboration between stakeholders and interoperation among heterogeneous applications

INDUSTRY INVOLVEMENT

We expect that our work will be of interest to many organizations from industry and government agencies. During the first 6-months of the first year seed project, we have established strong collaborative working relationship with Prof. Hans Bjornsson of Chalmers University (and UC Merced) on the "ISCIS – Integrated Supply Chain Information Systems" project jointly conducted at VTT (Technical Research Centre of Finland) and Chalmers. NIST has identified information interoperability and supply chain management as one of the key initiatives. We have interacted and discussed our research with Dr. Robert Lipman (on IAI's ifc and CIS/2 Interoperability project) and Dr. Mark Palmer (on FIATECH's AEX Interoperability project) in NIST's Building Integrated Building Process Group. Additionally, we have established a collaborative research effort with Dr. Albert Jones of the Enterprise Systems Group at NIST, who is currently providing partial funding support for Dr. Jiayi Pan, a post-doc student, on a related topic of metrology interoperability. Additionally, we plan to work closely with Mr. Dave Conover of the International Code Council towards the development of next generation "smart" code (a Building Smart Initiative) which involves interoperability between building codes and models. Dr. Calvin Kam of GSA, who has significant experience in interoperability standards and building models, has expressed interest in our research. We expect these collaborations will continue in this proposed project towards establishing synergistic research activities to improve the development of interoperability standards for the AEC industry. Our plan will include presenting our results and findings at NIST's and FIATECH's workshops to solicit feedback and suggestions. Last but not least, we anticipate feedbacks and participations from facility managers and owners, government agencies and CIFE company members.

RESEARCH PLAN, SCHEDULE AND RISKS

Organization and key personnel

This proposed interdisciplinary research involves investigators from Civil and Facility Engineering (Prof. Kincho Law and Dr. Chuck Han) and Computer Science (Prof. Gio Wiederhold). Prof. Wiederhold has been involved in information science research for over 40 years and has long envisioned the important role of mediation for information systems and web services. Dr. Chuck Han (Consulting Assistant Professor) is an expert in ICT research and practice in CAD and facility engineering. Prof. Law has been actively involved in engineering information management and enterprise integration for over 20 years. The research team will also be assisted by an assistant, Mr. Jack Cheng, whose PhD research focuses on information interoperability and ICT in Supply Chain Management. Researchers from Chalmers, NIST and other organizations will also be actively participating in this research project.

Research Tasks, Schedule and Milestone

We anticipate the following results from the proposed research:

- By the end of Autumn Quarter – We plan to develop methodologies for the mediation tools, and establish information interoperability infrastructure.
- By the end of Winter Quarter – We plan to implement and test the information interoperability result on a simple procurement supply chain application
- By the end of Spring Quarter – We plan to define and scope a series of case scenarios with increasing complexity jointly with collaborators (such as NIST and GSA) to demonstrate the interoperability among different applications
- By the end of Summer Quarter – We plan to select a set of case scenarios of modest scale (a building project with 3-5 disciplinary domains) to demonstrate the methodology and solicit feedback from internal and external use.

Validation and calibration of the interoperability methodologies will be an ongoing activity, which we do not expect to complete by the end of the project calendar year. The purpose is to identify the needs and formulate the theoretical framework. Further research tasks will be the focus of subsequent years (with the expectation of funding from outside sources).

Risks

We recognize that to establish a comprehensive information system infrastructure to support information interoperability and web service integration of a construction supply chain is a task beyond a 1- or 2-year seed research project. However, pilot projects such as the one proposed here represents a feasible objective that can lead to fundamental understanding of the research and development effort needed for broader implementations. We plan to leverage the collaborations with other researchers and their organization to maximize the output of this proposed research. This research will address some critical interoperability questions faced in our industry: What types of interface standards, modeling tools, and test methods will be needed to capture and exchange the semantics in construction supply chain? What types of standards, tools and methods are needed to deal with today's and tomorrow's technology? Which emerging standards and methodologies will provide the needed functionality for new advanced applications? We firmly believe this proposed seed research will contribute to these critical issues. The collaboration among the researchers from the different background and expertise will help achieve our objectives and goals.

NEXT STEPS

We plan to continue this research by exploring government funding opportunities such as the National Science Foundation, NIST and others. As noted, NIST (as well as FIATECH) has identified information interoperability and supply chain management as one of the key initiatives in their organization and has proposed for significant funding and supports in this area. We have already received partial funding support for a post-doc student from NIST on a related topic. With initial results from this proposed seed research project, we anticipate further funding supports from NIST as well as from NSF on this critical problem. This research represents a critical area of ICT in global construction and enterprise integration. We will also search for collaborative and funding opportunities with international partners.

References

Andrews, T., et.al. (2003), *Specification: Business Process Execution Language for Web Services (BPEL4WS)*,., IBM.

Ankolekar A, et al. (2001), "DAML-S: Semantic Markup for Web services," *Proceedings of the International Semantic Web Working Symposium,* Stanford, CA, pp. 411-430.

ANSI ASC X12C Communications and Controls Subcommittee. (2002), *ASC X12 Reference Model for XML Design*, Technical Report Type II ASC X12C/TG3/2002-xxx.

Begley, E. F., Palmer, M. E. and Reed, K. A. (2005), *Semantic Mapping Between IAI ifcXML and FIATECH AEX Models for Centrifugal Pumps*, Technical Report NISTIR 7223, NIST.

Booth, D., and Liu, C., (Eds) (2004), *Web Services Description Language (WSDL)*, W3C.

Cheng, J. Trivedi, P. and Law, K. H.. (2002) "Ontology Mapping Between PSL and XML-Based Standards For Project Scheduling," *Proceedings of 3rd International Conference on Concurrent Engineering in Construction*, pp. 143-156, Berkeley, CA.

Cheng, J., Cheng, C. P. , Thi, M. A. L. and Law, K. H.. (2006) “Engineering Simulations with Web-based Services,” *The 11th International Conference on Computing in Civil and Building Engineering, (ICCCBE XI)*, Montreal, Canada.

Coleman, G. and Jun, J. (2005), *Interoperability and the Construction Process – A White Paper for Building Owners and Project Decision Makers,* AISC.

Construction Specifications Institute (CSI). (2006), *OmniClass Construction Classification System*, Edition 1.0, http://www.omniclass.org.

Crowley, A., Watson, A. (2000), *CIMsteel Integration Standards Release 2*, SCI-P-268, the Steel Construction Institute, Berkshire, England.

de Bruijn, J., Martin-Recuerda, F., Manov, D. and Ehrig, M. (2004), *State-of-the-art Survey on Ontology Merging and Aligning* V1.SEKT-project report D4.2.1 (WP4), IST-2003-506826, EU-IST Integrated Project (IP), EU.

de Bruijn, J., et al. (2005), *Web Service Modeling Ontology (WSMO)*, Technical Report, DERI.

Euzenat, J., Le Bach, T., Barasa, J., et.al. (2004), *State of the Art on Ontology Alignment*, Technical Report KWEB/2004/D2.2.3/v1.2, EU-IST Knowledge Web (KWEB), EU.

Ehrig, M., and Sure, Y. (2004), “Ontology Mapping – An Integrated Approach,” *Proc. of the First European Semantic Web Symposium*, Volume 3053, Lecture Notes in Comp. Science, pages 76-91.

Gallaher, M., O’Connor, A., Bettbarn Jr., J., and Gilday, L. (2004), *Cost Analysis of Inadequate Interoperability in the US Capital Facilities Industry*, Technical Report GCR 04-867, NIST.

Gruninger, M., and Menzel, C. (2003), “Process Specification Language: Principles and Applications,” *AI Magazine*, 24:63-74,.

International Alliance for Interoperability (IAI). (1997), *Guidelines for the development of industry foundation classes*, IAI.

International Conference of Building Officials (ICBO). (2006), *International Building Code (IBC) 2006*, Whittier, CA.

Jacobs, J. and Linden (2002), A. *Semantic Web Technologies Take Middleware to the Next Level*, Technical Report T-17-5338, Gartner Group.

Lee, H.L., So K.C. and Tang C.S.. (2000). “The value of information sharing in a two-level supply chain,” *Management Science*, Vol. 46, No. 5.

Lee, H. L. and Whang, S. (2000), “Information Sharing in a Supply Chain,” *Int. J. of Manufacturing Technology and Management*, vol. No.1, pp. 79-93.

Leymann, F. (2001), *Web Services Flow Language (WSFL)*, IBM Corporation

Lipman, R. R. (2006), “Mapping Between the Cimsteel Integration Standards and Industry Foundation Classes Product Models for Structural Steel,” *Proc. of Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, Montreal, Canada, pp. 3087-3096.

NIST (1999), *Interoperability Cost Analysis of the US Automotive Supply Chain*, Planning Report 99-1, NIST Strategic Planning and Economic Assessment Office.

Preist, C. (2004), “A Conceptual Architecture for Semantic Web Services,” *Proceedings of the Third International Semantic Web Conference*, pp. 395-409.

Ray, S. (2002), “Interoperability Standards in the Semantic Web,” *Journal of Computing and Information Science in Engineering*, ASME, Volume 2, pages 65-69.

Teague, T. L., Palmer, M. E. and Jackson, R. H. F. (2003), “XML for Capital Facilities,” *Leadership and Management in Engineering*, Vol. 3, Issue 2, pp. 82-85.

van Hage, W., Katrenko, S., and Schreiber, G. (2005), “A Method to Combine Linguistic Ontology-Mapping Techniques,” *Fourth International Semantic Web Conference (ISWC),* pages 732-744.

Wiederhold, G. (1994), “Interoperation, Mediation, and Ontology’s,” *Fifth Generation Computer System, Workshop on Heterogeneous Cooperative Knowledge-Bases*, W3:33-48, 1994.