STANDARD DATA MODEL FOR CUSTOMS EDI FILINGS

Michael L. Grecol
Georgia Southern University
mgrecol@hotmail.com

ABSTRACT

Electronic Data Interchange (EDI) is a popular way in which business entities exchange transactional data. A significant challenge in using EDI is that different entities have different specifications for the transmission of the same data. An example of this is border security filings that are required by many countries in order to gain admissibility of imported merchandise. Various countries require different EDI data formats. This paper defines an integrated data model, the Single Filing System (SFS) as a standard, applicable to containerized sea freight shipments to Canada and the U.S., and an industrial-strength proof of concept for the model integration methodology. It simplifies the processing of shipment data for multiple heterogeneous EDI data formats prior to exporting containerized sea freight.

Keywords

EDI, B2B protocol, data integration, data model, data formatting, customs, security filing, ABI, single-window, SAFE Framework

BACKGROUND

Businesses and governments use EDI to share information; however, because of the multitude of standards available, they do not always agree on the same EDI standards. This point is illustrated in the implementation of recent business to customs EDI filing requirements. Since the terrorist events of September 11, 2001, "technologically advanced and unburdensome procedures for inspecting and clearing cargo (Peterson and Treat, 2008)" were needed in order to secure the supply chain. The solution was to require the shipping parties to file electronic cargo information in advance of the shipment exportation through EDI. The World Customs Organization (WCO) introduced the SAFE Framework and Single-Window and the United Nations introduced UNCEFACT to standardize these EDI formats (Ireland, 2009; UNECE, 2011; WCO, 2007). However, the U.S.A.'s Importer Security Filing, did not conform to any standard messaging format required by the SAFE Framework nor with the UN data standards (Blegen, 2008; UN, 2005; UNECE, 2011). The U.S. ISF EDI is incompatible with other EDI formats (Blegen, 2008; USCBP, 2010; USCBP, 2011). This paper proposes an integrated data model named the Single Filing System (SFS) as a standard that captures the variability of EDI inputs and outputs providing proof of concept for model integration methodology.

DESIGN PROCESS OF THE SINGLE FILING SYSTEM

This paper proposes the SFS standard by using the example of containerized sea freight shipments to Canada and the U.S. because the filings are similar but with some significant structural differences. To arrive at a standard model, the following design process was used: 1) model the target EDI specifications, 2) model the semantics of the logistics ontology, and 3) employ a design approach to arrive at a data model that can be used to drive several different outputs.

MODELING THE U.S. ISF FROM THE EDI SPECIFICATION

Modeling from the EDI specification is the first step to understanding the data requirements. Since the only source of the dataset design is the EDI specification provided by U.S. Customs, the Entity-Relationship diagramming technique known as Identification Technique, particularly output data analysis structuring, was used while maintaining a Data-Oriented Approach throughout the process (Jovanovic, Gardiner and Cupic, 2007; Ponniah, 2007).

The U.S. ISF uses the proprietary format known as the Automated Broker Interface or ABI format (Schulz and Orlowska, 2001). An analysis of the EDI specification along with several data samples was conducted to arrive at the simplified model in Figure 1. The SF20, SF25 and SF31 records are optional (USCBP, 2011). The SF30-36 records must be transmitted with information of the manufacturer, seller, buyer, ship to location, container stuffing location and consolidator. It is required to send at least one, SF40 Harmonized Tariff Schedule Record with the last SF30-SF36 records (USCBP, 2011). The Harmonized Tariff Schedule is the internationally accepted codes used to classify commodities (WCO, 2011). Also note that the only required shipment information is the SF15 which contains the Bill of Lading Number (USCBP, 2011).

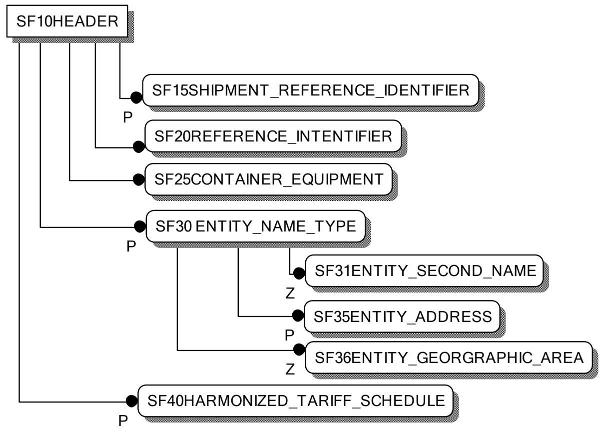

Figure 1. U.S. ISF EDI Data Model

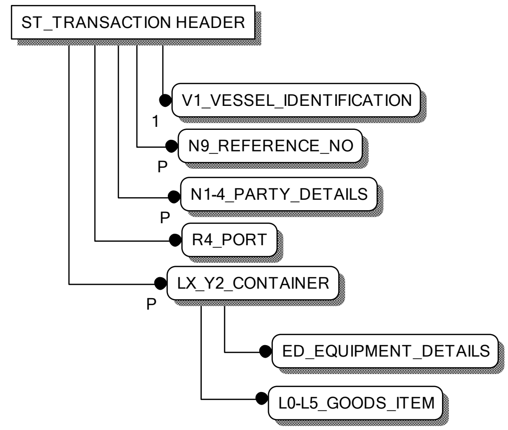

Figure 2. Canada ACI EDI Data Model

MODELING CANADIAN ACI FROM THE EDI SPECIFICATION

The Canadian ACI filing is much like the U.S. Customs ISF Filing; however, there are differences with transmission format as well as the relationship between data elements. The Canadian government accepts both ANSI and EDIFACT formats. Due to the simplicity, the ANSI model has been modeled in Figure 2 (CBSA, 2008; CBSA, 2008a). The N1-4 record must be sent with information about the shipper, consignee and delivery address. Shipping port information is contained in the R4 record and the Harmonized Tariff Schedule code is located within the L0-L5 records which has a one-many relationship with the LX-Y2 Container record (CBSA, 2008a).

SEMANTICS OF THE MARITIME LOGISTICS ONTOLOGY

It is accepted that an analysis of the semantics of an ontology helps to arrive at a suitable standard model (Hull and King, 1987; Nie, Fu and Xia, 2009). Bichou et al has developed a conceptual representation of an international maritime network shown in Figure 3 (Bichou, Lai, Lun, and Cheng, 2007). This representation has three channels and three flows (payment, information, and physical). For the purpose of this discussion we are concerned with both the information flows and physical flows.

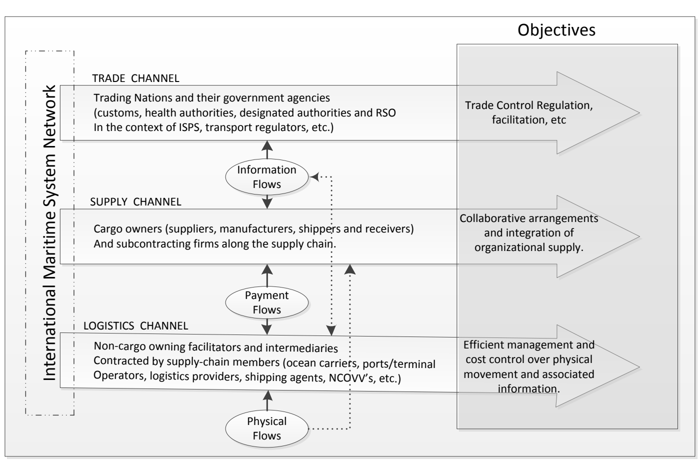

Figure 3. International Maritime Network (Bichou, et al., 2007)

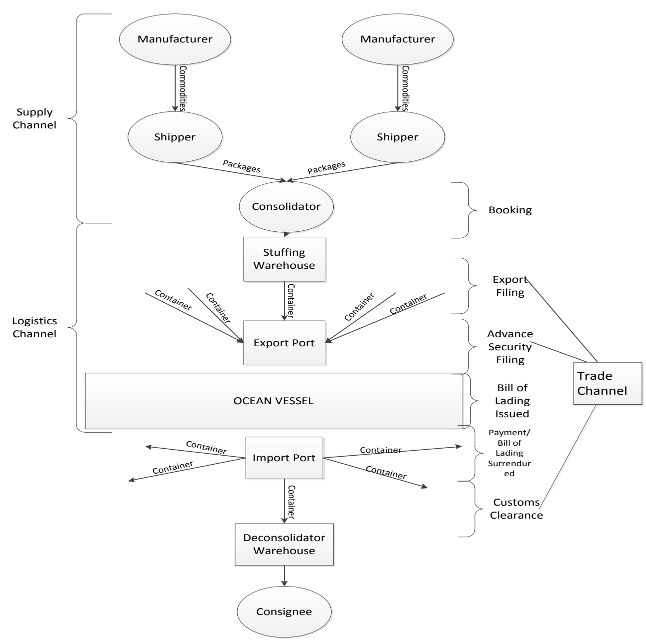

Figure 4. Container Logistics Workflow Diagram

The security filings are part of the trade channel. The dotted line named "information flows" from the supply channel to the logistics and supply channels in Figure 3 represent the security filings and other trade data. These filings should represent the semantics of both the supply and logistics channel ontologies.

Business workflow diagrams help to ensure that users and standards developers have a common understanding of the domain. (Huemer, 2001). The container logistics workflow diagram is shown in Figure 4.

The diagram represents the following facts of maritime container logistics. A booking is issued for one or many containers. One or more packages containing one or more commodities are sent to a stuffing warehouse for stuffing into the ocean container. Once stuffed and sealed, an export filing is made to the export country customs agency, then the container is transported to the export port. The export port receives containers for one or more vessels.

Before vessel departure from the export country, an advance security filing is transmitted (CBSA, 2008; USCBP, 2010). The carrier issues the bill of lading which can be for one or many containers after the containers are verified on board and the vessel departs. These processes will be the main focus of this paper's proposal.

The crux of modeling maritime logistics depends upon achieving a close match between ontological semantics and the data model. Neither the ISF nor the ACI data models represent these semantics accurately. For example, the ISF relates commodities directly to the header, while the ACI relates commodities to the container record (CBSA, 2008; USCBP, 2011).

DESIGN OF SFS

The challenge of the SFS design is the method of data integration. Bussler reviews three integration approaches in his RIDE'02 conference proceedings (Bussler, 2002).

1) Programming-Based Integration Approach.
2) Naïve Modeling Approach
3) Advanced Modeling Approach

Programming-Based Integration Approach uses a programming language like Java to extract messages from backend systems (Bussler, 2002). This approach is not recommended due to a high number of messaging combinations. Each message format would need a separate program; therefore, a large number of programs would have to be developed and maintained to support the various protocols (Bussler, 2002).

The Naïve Modeling Approach known as a workflow type approach models workflow types. It is not recommended because it does not model sequences or transformations. Furthermore, the workflow system's database refers to a set of programs with similar limitations as the above (Bussler, 2002).

The Advanced Modeling Approach has the advantage that every aspect of the business transaction is modeled from protocol, sequencing and workflow types (Bussler, 2002). Bussler suggests the process of binding business documents to a B2B protocol much like the Business Process Broker described by (Bussler, 2002; Schulz and Orlowska, 2001). The binding layer transforms data to a normalized format which can be more easily processed into database application (Bussler, 2002; Kuchibhotla, Dunn and Brown, 2009). Using this methodology, the SFS database can be a generic database representing semantic logistics and e-commerce data. Binding this data to the B2B protocol will produce the acceptable electronic documents required by foreign governments. This means that the SFS system will consist of a model for the standard document data and a model for the B2B protocol allowing a homogeneous database to drive heterogeneous data outputs.

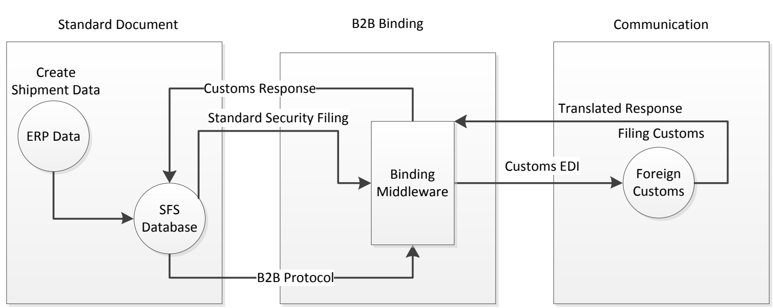

Figure 5. SFS Design Concept

This paper proposes that the SFS model, illustrated in Figure 5, shall be based on Bussler's Advanced Modeling Approach. The B2B protocol shall be stored in the RDBMS which is retrieved by an application written in Java or similar language to produce a standard document type from the database while the B2B protocol creates a transformation tool as such described in (Bussler, 2001).

The final SFS database model shown in Figure 6 represents the semantics of the logistics and supply channel ontologies. The entities on top of the model represent the standard document format while the bottom entities represent the B2B Binding model. The B2B Binding model allows the middleware to convert the normalized data document to the specified country's EDI format. The conversion is model based, and does not rely on programming changes to facilitate any change in filing format. It is a proof of concept of a normalized data document being converted to various specific formats through application middleware as described by (Bussler, 2002; Zheng, 2009). Below are some details on the specific entities in the SFS model.

The Party entity contains the name and location of every party in the transaction. The Customs_Party_ID records represent country-specific codes assigned to the parties in the transaction. Validity is tested by joining to the Customs_ID_Codes entity and testing the code against the Regular Expression identifiers (The Open Group, 1997). The central entity in the standard document is the Bill_of_Lading entity. The remainder of entities is related to the Bill_of_Lading entity either directly or indirectly. The Bill_of_Lading can have one or many Invoice records. Each Invoice record must have one or many Invoice_Part_Detail records. The Harmonized_Tariff_Code entity is related to each Part_No in Invoice_Part_Detail based on the effective date. Having an effective date is important as the Harmonized_Tariff_Code record is a key component of compliance and may need to be changed periodically while maintaining a historical record of its assignment.

The B2B Protocol entities are used by the middleware to convert the standard document format to the EDI text. The Shipment and Shipment_Contents entities are used to identify a group of Bill_of_Lading records to be included in a single filing. The Filing entity contains the information about the target country and format to be transmitted in addition to derived information about the filing status. The Filing_Documents entity defines items that are common to the entire filing such as the delimiters, end of line identifiers, connection information and a stored procedure to retrieve the data matching the target output schema. The Filing_Line_Spec entity contains the details about each line of the EDI spec. For each line, one or many Filing_Spec_Item records must be populated. The Start_Position, End_Position and Fill_Char fields specify significant formatting required for positional EDI while the Start_String and End_String fields are used to specify any special identifiers or default data required to appear directly before or after the field data. Because of differences between standardized codes among filing systems, a conversion table named String_Conversion is used by the binding layer to convert these elements to the target code standard. The To_Customs and From_Customs entities contain the EDI text of the outgoing and incoming EDI messages respectively.

The steps to creating an EDI document from the standard document format are as follows:

1) Inserting a Shipment record and associating Bill_of_Lading records with it.
2) Retrieving an XML document based on a country specific XML Schema by using the stored procedure defined in the Filing_Documents table.
3) Middleware formatting to textual EDI using format details specified in the Filing_Line_Spec and Filing_Item_Spec tables.
4) Transmitting the formatted EDI to the appropriate location defined in the Filing_Documents table.

The SFS system supports reuse and builds upon the W3C standard through the use of XML schemas with frequently used entities defined as complexTypes, allowing them to be used by reference (Bean, 2003).

CONCLUSION

The SFS offers the functionality of multiple systems with the simplicity of a single model-driven standardized database. The alternative to the SFS may be either using separate systems for each country filing or employing custom programming to output each EDI format. The SFS is a vast improvement over these methods, resulting in efficiency and simplicity in complying with multiple country regulations.

ACKNOWLEDGMENTS

My sincerest gratitude is extended to Anggita Prawiranata and Dr. Vladan Jovanovic for comprehensive editing and invaluable improvement suggestions. In addition, thank you to the following who have provided assistance with various country data requirements: Satya Prasad Sahu (WCO), Nathalie Jubainville (CBSA), Jill Balfour (CBP), Vincent Annunziato (CBP).

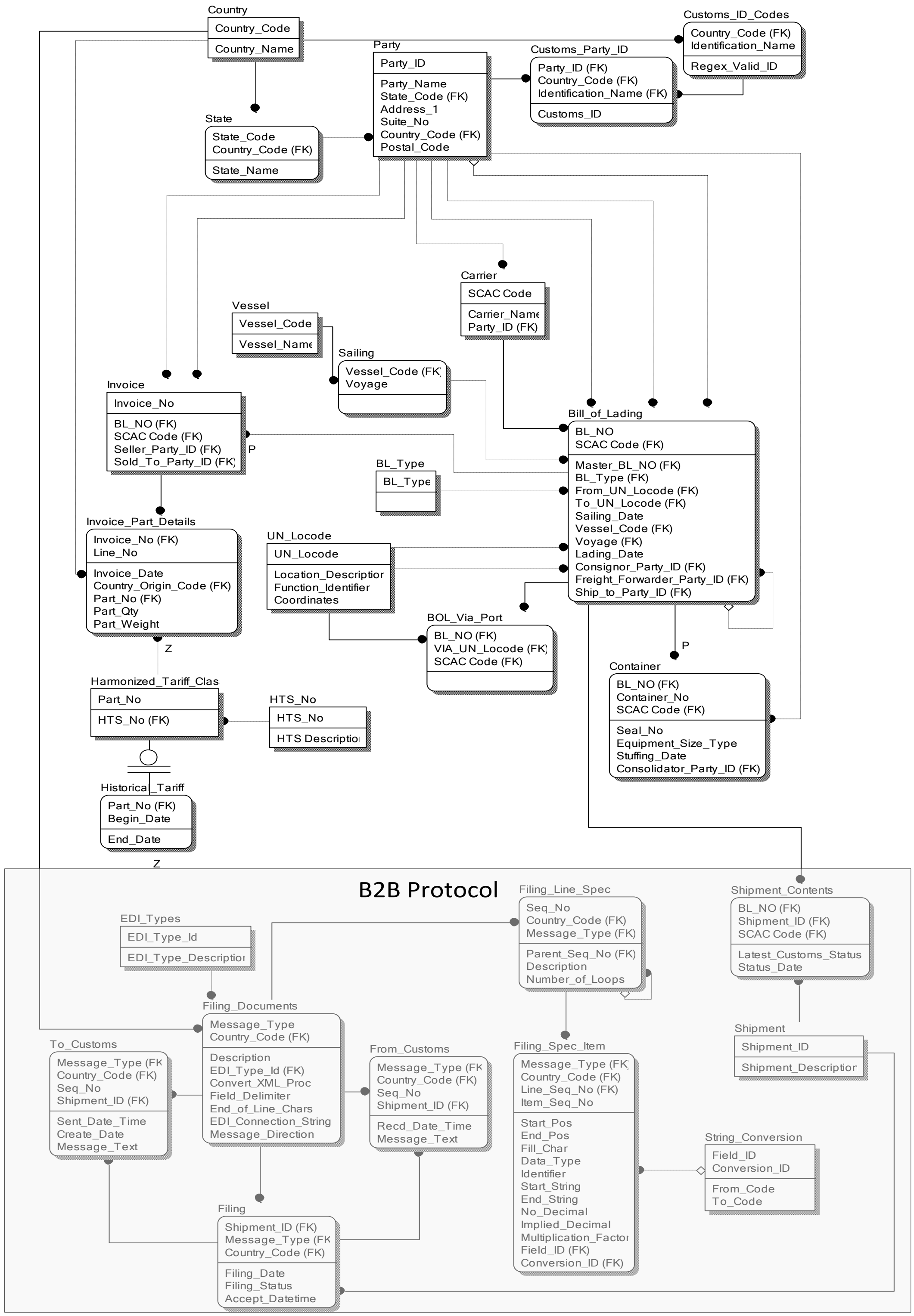

Figure 6. SFS Standardized Database Model.

REFERENCES

1. Bean, J. (2003) XML for data architects: Designing for reuse and integration, Morgan Kaufmann Publishers, Waltham, MA, U.S.A.
2. Bichou, K., Lai, K., Lun, Y. and Cheng T. (2007) A quality management framework for liner shipping companies to implement the 24-hour advance vessel manifest rule, *Transportation Journal*, 46,1, 5-21.
3. Blegen, B. (2008) US importer security filing: *Advance Electronic Data Under the SAFE Framework Meets the Real World*, *World Customs Journal*,3,1,73-85.
4. Bussler, C. (2001) Semantic B2B integration server technology as infrastructure for electronic hubs, *12th International Workshop on Database and Expert Systems Application*.
5. Bussler, C. (2002) Modeling and executing semantic B2B integration. *Proceedings of the 12th Int'l Workshop on Research Issues in Data Engineering.*
6. Canadian Border Services Agency (CBSA) (2008) Advance commercial information (ACI) Marine Client Document.
7. Canadian Border Services Agency (CBSA) (2008a) Memorandum D3-5-2.
8. Huemer, C. (2001) Defining electronic data interchange transactions with UML, *Proceedings of the 34th Hawaii International Conference on System Sciences.*
9. Hull R. and King, R. (1987) Semantic database modeling: Survey, applications, and research issues, *ACM Computing Surveys, 19*, 3.
10. Ireland, R. (2009) The WCO SAFE Framework of standards: Avoiding excess in global supply chain security policy, *WCO Research Paper No. 3.*
11. Jovanovic, V., Gardiner, A. and Cupic, L. (2007) A taxonomy of data modeling techniques, *Proceedings of the 2007 Southern Association for Information Systems Conference*, 189-194.
12. Kuchibhotla, H., Dunn, D. and Brown, D. (2009) Data integration issues in IT organizations and a need to map different data formats to store them in relational databases, *41st Southeastern Symposium on System Theory*, Tullahoma, TN.
13. Nie, G., Fu, M. and Xia, H. (2009) A semantic mapping system based on E-commerce logistics ontology, *World Congress on Software Engineering*, 133-136.
14. The Open Group (1997) Regular expressions, The single Unix specification Version 2, Retrieved from: http://pubs.opengroup.org/onlinepubs/007908799/xbd/re.html
15. Peterson, J. and Treat, A. (2008) Post-9/11 Global framework for cargo security. United States International Trade Commission, *Journal of International Commerce and Economics.*
16. Ponniah, P. (2007) Data modeling fundamentals: A practical guide for IT professionals, John Wiley & Sons, Inc., Hobeken, New Jersey, US.
17. Schulz, K. and Orlowska, M. (2001) Architectural issues for cross organisational B2B interactions, *21st International Conference on Distributed Computing Systems Workshops*, Queensland, Australia.
18. The United Nations (UN) (2005) Recommendation and guidelines on establishing a single window, *United Nations Centre for Trade Facilitation and Electronic Business*, New York and Geneva.
19. United Nations Economic Commission for Europe (UNECE) (2011) Retrieved from UNECE United Nations Economic Commision for Europe: http://www.unece.org/cefact/
20. U.S. Customs and Border Protection (USCBP) (2010) Importer security filing and additional carrier requirements: "10+2" Program Update, retrieved from: http://www.cbp.gov/xp/cgov/trade/cargo_security/carriers/security_filing/
21. U.S. Customs and Border Protection (USCBP) (2011) Customs and trade automation interface requirements, retrieved from: http://www.cbp.gov/xp/cgov/trade/automated/automated_systems/abi/catair/
22. World Customs Organization (WCO) (2007) WCO data model, single window data harmonization, World Customs Organization, Brussels Belgium.
23. World Customs Organization (WCO) (2011) Retrieved from World Customs Organization: http://www.wcoomd.org/home.htm.
24. Zheng, O. (2009) Application of XML-based heterogeneous database data exchange middleware in e-commerce, *2009 IITA International Conference on Services Science, Management and Engineering*, 132-135.