International Journal of Information Science and Management

'Push' and 'Pull' Modeling for RosettaNet's PIPs

Ting Tin Tin, Ph.D. Student
School of Computer Sciences
University of Sains Malaysia
email: tingtintin@yahoo.com

Vincent Khoo, Ph.D.
School of Computer Sciences
University of Sains Malaysia
email: vkhoo@cs.usm.my

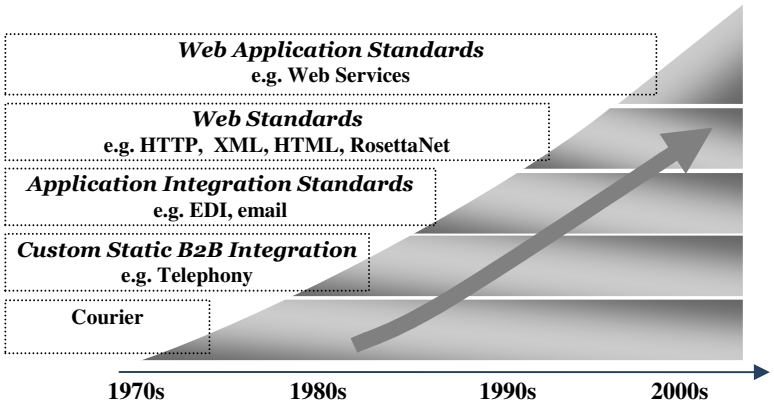
Abstract

This paper analyses and compares the performances of Business-to-Business e-Commerce data quality and service quality among the use of the 'Push' and 'Pull' models. RosettaNet's original Partner Interface Processes use the 'Push' model which suffers from data redundancy when the volume of business-to-business transaction increases. However, when the 'Pull' model is used, although it reduces data redundancy, it has a greater chance of not getting the latest updates. It is then argued that if the 'Push' and 'Pull' models are combined, the resulting model would further enhance the overall trade document interchange performance with improved data quality and greater personalizability, especially among the smaller non-Electronics industries.

Keywords: Push Model, Pull Model, RosettaNet Partner Interface Processes, Data Quality and Service Quality.

Introduction

Information and Communications Technology (ICT) helps organizations achieve high performance in information access from shared data banks and the Internet (Accenture, 2004). One of the electronic business applications in the ICT industry is electronic commerce which uses electronic communications technology in the daily business transactions (Wikipedia, 2007a). Since 1990s, many organizations have been moving towards electronic business document interchange (RosettaNet, 2002), as shown in Figure 1. The emergence of the Internet has encouraged the evolution of electronic document interchange from traditional electronic data interchange (EDI) into Internet-based data interchange. However, the high initial investment cost has slowed down the adoption of EDI (Wikipedia, 2007b), especially among the small and medium-sized industries.



Adapted from RosettaNet Executive Insights (RosettaNet, 2007b) & Beyond e-Marketplace & Next Generation e-Business (Chung, 2002).

Figure 1. Business-to-business technologies evolution.

This paper represents the initial work of a series of projects described in Ting, Khoo & Cheah (2007). The main research objective of Ting et al. (2007) is to model, analyze, and design a next generation Business-to-Business Standards Component Model in a service oriented architecture by using Web services on an open platform for a more effective interchange or even sharing of trade documents in a more personalized manner.

RosettaNet is a non-profit organization established in 1998 defining standards in global supply chain. RosettaNet Partner Interface Processes (PIPs) standardize business processes by defining business documents format and content in seven sectors: demand creation, design, forecast, order, payment, logistics, and manufacture. Each document is complemented with a business document specification, business process specification, PIP process specification and message structure to ease the implementation of the standards. Through the standardized business processes and document content and format, the flow of data can be easily observed, analyzed and controlled to improve the overall efficiency of the business-to-business (B2B) integration. RosettaNet standards have been widely endorsed by some 500 companies worldwide. More details about RosettaNet PIPs standards can be found in (RosettaNet, 2007a).

Since RosettaNet has been used by the project partners, its trade document interchange standards have first been chosen in this research project. RosettaNet standards increase the efficiency of business processes by reducing the number of manual transactions, contract costs, inventory and administrative costs, shipments and logistics costs, etc. (RosettaNet, 2007d). Moreover, the standards facilitate direct connections between trading partners without any third-party intermediary such as Value Added Networks (Damodaran, 2005). Over the years, RosettaNet has reduced the costs of Partner Interface Processes creation. The cost improvements include XML

schema representation, shared UML models, and automated XML schema creation. RosettaNet Automation Enablement (RAE) is another program introduced to help the small and medium-sized industries to adopt the RosettaNet standards. RAE eliminates RosettaNet Implementation Framework development which enables the 24/7 Internet connection among the smaller industries. However, there are problems in maintaining data quality and service quality when the volume of Business-to-Business transaction increases (Damodaran, 2004). The original Partner Interface Processes use a 'Push' model which has a high rate of data redundancy (Damodaran, 2004). The size of a Partner Interface Processes is usually between ten megabytes and a hundred megabytes. A complete Partner Interface Processes business transaction normally consumes hundreds of gigabytes. As a result, the need for larger disk storage and higher network speed will increase exponentially as a business grows. Therefore, a 'Pull' model is introduced to pull only the required data. Nevertheless, the 'Pull' model has its drawback of not getting the most current data (Bhide et al., 2002). Based on the research done by Bhide et al. (2002) which describes the dynamic dissemination of Web data, a combination of the 'Push' and 'Pull' models is proposed in this paper for trade document interchange among the users' trading partners.

RosettaNet Partner Interface Processes (PIPs)

A case study has been carried out in Schönberger (2006) to realize the RosettaNet PIPs compositions as Web service orchestrations. It proposes a framework to execute PIPs using a 'Push' model through some Web services. While the public business processes have been extensively standardized in the document format and interchanged among the trading partners using the 'Push' model, this research project attempts to personalize some of the private processes through Web services in some other models such as the 'Pull' Model. Prior to this project, the same research team has reviewed and analyzed some possible approaches to enhance B2B integration in the RosettaNet environment. The research outcome has been documented in the form of a research roadmap in (Ting et al., 2007).

B2B procurement process using PIPs

This section examines the 'Push' model by observing the business documents interchanged in a typical B2B procurement process as shown in Figure 2. Based on the latest update from the RosettaNet Website when this paper was written, there was a total of 109 PIPs to-date (RosettaNet, 2007a). Figure 3a and 3b show how the PIPs and RosettaNet Implementation Framework (RNIF) function in a B2B integration process.

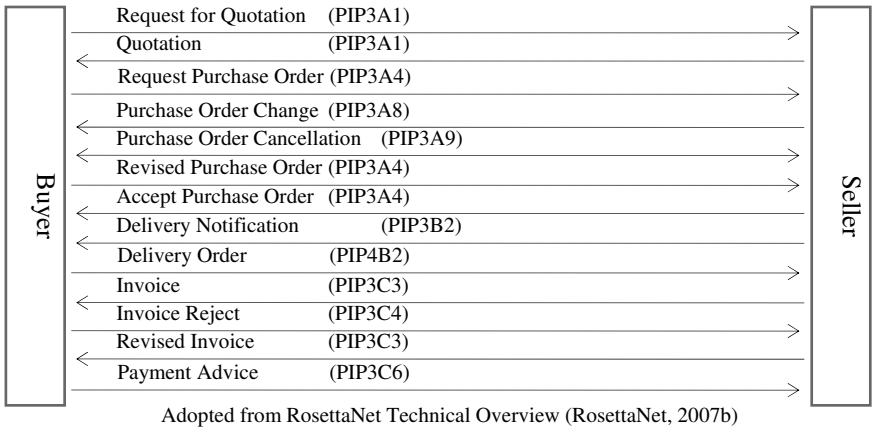


Figure 2. B2B transactions pushed through PIPs.

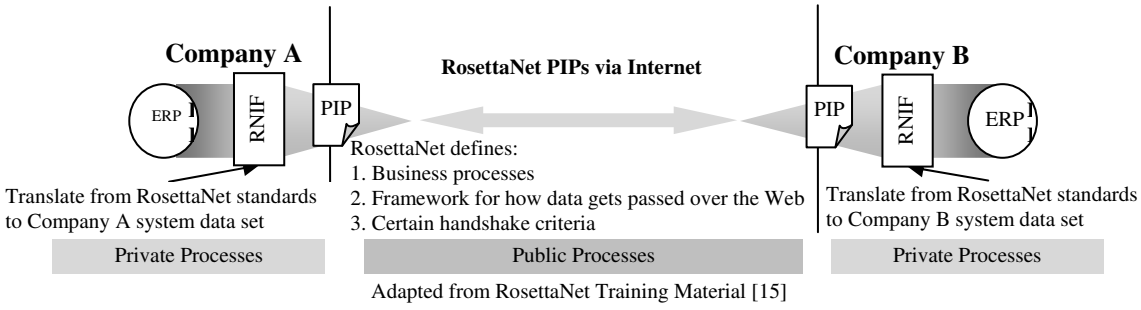


Figure 3a. PIPs and RNIF in B2B integration.

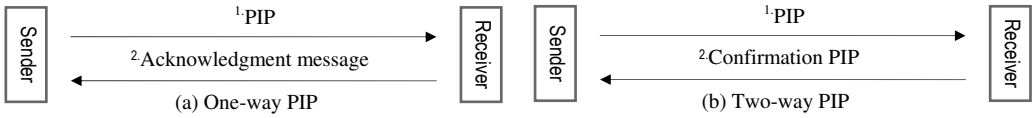


Figure 3b. One-way PIP and two-way PIP.

Emailing is a good example of using a 'Push' model for document interchange. A sender can send an unlimited number of emails to a receiver without the need for any response. As a result, junk messages tend to overload the mailboxes.

Data quality and service quality in PIPs

The multi-national companies in the Electronics industry use PIPs and RNIF to interchange business documents with their trading partners. The sequence of trade document interchange is predefined in the Trading Partner Agreement. Based on the data provided by a large international electronic company (Intel), more than 90% of its trading partners are the small and medium-sized industries. As such, it is necessary to study and analyze the data quality and service quality of PIPs so that the small and medium-sized industries will be more ready to adopt the standards.

Data quality

Relevancy and completeness are two of the main attributes for data quality. Enterprises invest heavily to provide customized documents in some specific format and content to meet the needs of their partners' requests on data relevance and completeness. Although PIPs take into account the entire content of trade documents which comply with completeness, Damodaran (2004) asserts that PIPs have a problem in 'large messages with redundant content'. In this study, an analysis is carried out to compare the contents of two PIPs, they are Request Purchase Order and Query Order Status. Among the first 100 data fields in Query Order Status, some 90 data fields have been found to be redundant in the Request Purchase Order. In other words, some 90% of the information exchanged in these two processes are redundant. The entire Request Purchase Order and Query Order Status have accumulated up to some 50% of redundant fields. Therefore, a large amount of redundant data has actually been interchanged among the partners. It is pointed out in Surfcontrol (2007) that large amount of emailing reduces network efficiency and speed and thereby affects a company's ability to meet tight deadlines. Likewise, redundant data could also clog the network and delay the business processes. Nevertheless, customizing or personalizing documents to meet different partners' request is very costly and time consuming. This particular need has motivated this research to provide personalization Web services for the small and medium-sized industries.

PIPs standardize the documents' format and content. In the business processes, many business documents need to be produced manually or system generated. For example, in PIP3A (Quote and Order Entry), at least ten business documents with thousands of data fields need to be created and sent to the trading partners. PIP Request Purchase Order and PIP Query Order Status are two of the primary business documents involved. As the business grows, the content of a predefined document needs to be updated. In PIPs, it is costly and time consuming to alter the content. Therefore, partners are reluctant to change the predefined content. In short, although PIPs standardize the document content, they sometimes do not comply with the 'relevancy' requirement of the users.

Service quality

In business-to-business document interchange, electronic document standardization has been introduced to ease the communications among the trading partners. From the perspective of the small and medium-sized industries, the standardized formats have increased data redundancy, communication overheads, storage overheads, as well as the inflexibility in satisfying their respective needs and wants. Three possible personalization opportunities are described in Hanson et al. (2007), including (a) the

products/services, (b) the Website where the partners interact, and (c) the communication and messaging that reach the partners through a variety of channels and media.

In addition, real time accessibility of data depends on the availability of the sender. If a small company does not push out the data, the receiver will never get it. Accessibility refers to a situation whether a request for service could be met. The 'Push' model has lower accessibility in terms of data retrieval (Bhide et al., 2002).

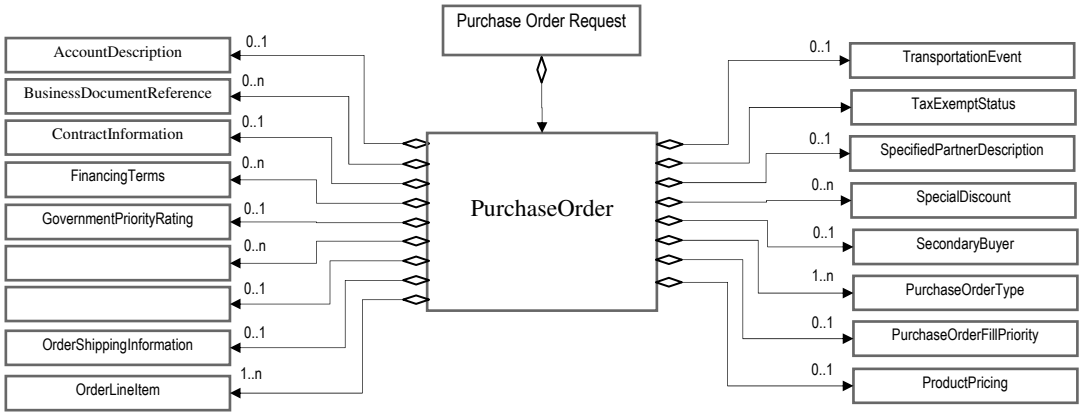
Alternative Models of B2B Document Interchange

In view of the drawbacks identified in the 'Push' model using the original PIPs, this paper attempts to propose other alternative models for enhancing the B2B document interchange through PIPs.

Document interchange using data-blocks

It is proposed to use a so-called 'data-block structure' for minimizing data redundancy. RosettaNet PIPs use the UML models to represent the business requirements. XML messages are produced based on the UML models. Inheritance and aggregation are used to achieve consistency and reduce the size of the XML messages Damodaran (2005). XML messages are produced based on the UML model shown in Figure 4.

The data redundancy stated in Damodaran (2004) can be identified and estimated from the above-mentioned UML models. The Purchase Order Request and Purchase Order Change Request are used for discussion, as shown in Figure 5. The available data groups for each PIP are represented in alphabets.



Adapted from RosettaNet PIP Directory (RosettaNet, 2007a)

Figure 4. UML model for Purchase Order Request.

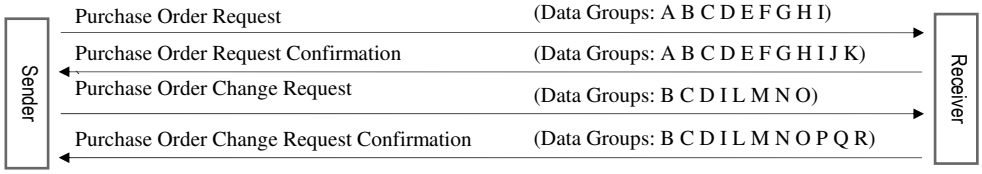


Figure 5. Four typical business documents interchanged.

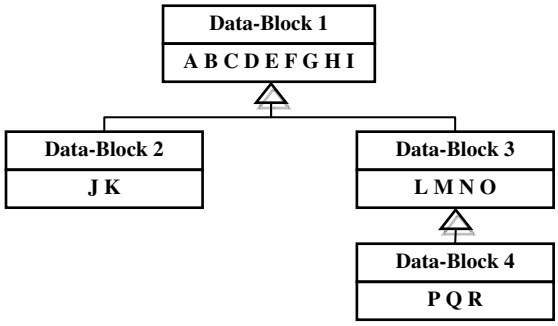


Figure 6. Possible data-block hierarchy.

In Figure 6, data-block 4 inherits all or some classes from data-block 1 or data-block 3. Therefore, data-block 4 may have classes such as C and D from data-block 1 and classes L and M from data-block 3. The nature of such inheritance is predefined in the Trading Partner Agreement. Figure 7 shows the document interchange using the proposed data-block hierarchy.

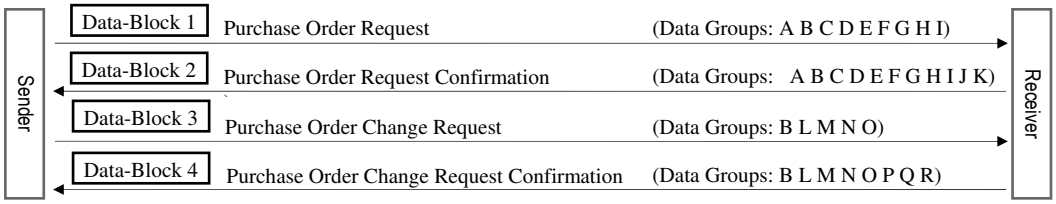


Figure 7. Data-blocks interchanged.

In a typical procurement process, data-blocks are interchanged accordingly. Data sent in the previous data-block will not be re-sent in the second data-block unless there are changes to the data. This reduces data redundancy. The data-block hierarchy also helps to ensure the completeness of data. The receiver can check the received data based on a checklist to ensure that no data elements are missing.

'Push' model using data-blocks

The data-block structure can be used to reduce redundancy in the original 'Push' model as shown in Figure 8. However, the main drawback of this model is that the receiver does not have the control over what and when to receive the data.

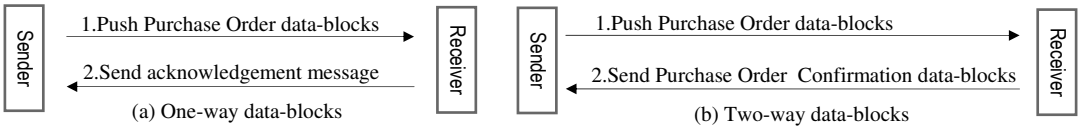


Figure 8. 'Push' model using data-blocks.

'Pull' model using data-blocks

In the receiver-oriented 'Pull' model, the receiver has the control over what and when to receive the data. However, the receiver might not be aware of the latest updates and therefore the fidelity is lower compared to the 'Push' model.

Figure 9. Receiver-oriented 'Pull' model.

'Push and Pull' model using data-blocks

The 'Push and Pull' model is introduced in this study to overcome the shortcomings arising from the previous models. There is an intermediary server between two trading partners. A sender pushes the updated data to the intermediary server with a confirmation message in return. A receiver can then pull the desired data from the intermediary server. Therefore, the fidelity for this model is high.

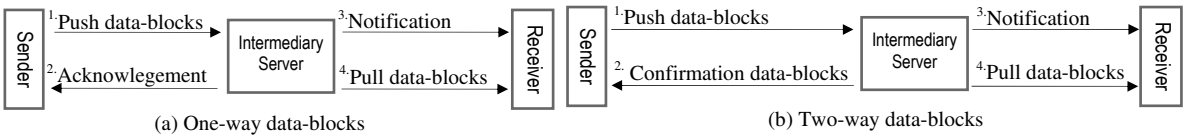


Figure 10. 'Push and Pull' model using data-blocks.

'Push and Pull' model using data-blocks

The 'Push and Pull' model is introduced in this study to overcome the shortcomings arising from the previous models. There is an intermediary server between two trading partners. A sender pushes the updated data to the intermediary server with a confirmation message in return. A receiver can then pull the desired data from the intermediary server. Therefore, the fidelity for this model is high.

'Push and Pull' model using Web services

In order to provide personalizable capabilities to the trading partners, Web services can be incorporated in the 'Push and Pull' model. The trading partners can pull any

desired data from the intermediary server upon receiving a notification. This is supposedly the most ideal model for the small and medium-sized industries to freely access the most up-to-date data anytime with minimal data redundancy. In effect, a decision support Web service could be implemented at the intermediary server through which the users can set the prioritized criteria and in return the Web service would recommend the most appropriate model(s) for the users to make more-informed document interchange decisions. Some of the criteria may include cost, accessibility, tolerable level of data redundancy, frequency of data updates, flexibility in timing and content of document interchange, etc

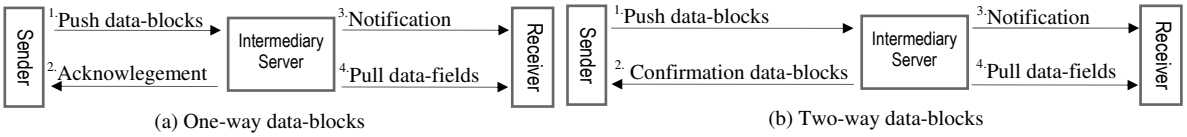


Figure 11. 'Push and Pull' model using Web services.

Performance of Document Interchange Models

The small and medium-sized industries have been slown in adopting the RosettaNet standards for document interchange mainly due to its high initial cost, which may be attributed to the high data redundancy and low accessibility in using the 'Push' model with the original PIP formats. Four different models were proposed in the previous section to encourage more widespread use of the RosettaNet standards. The models are evaluated in terms of their data quality and service quality, as shown in Table 1. The evaluation is based on four criteria namely, data redundancy, fidelity, accessibility and personalizability.

Table 1

Performance of Various Models

No.	Model	Data Quality		Service Quality		
		Data redundancy	Fidelity (Data updates)	Accessibility	Personalizability	
					Timing	Content
1	'Push' model using original PIPs	High	Updated	Low	Not flexible	Coarse grain
2	'Push' model using data-blocks	Low	Updated	Low	Not flexible	Coarse grain
3	'Pull' model using data-blocks	Low	Out-dated	Low/Medium	Flexible	Medium grain
4	'PnP' model using data-blocks	Low	Updated	High	Flexible	Medium grain
5	'PnP' model using Web services	Low	Updated	High	Flexible	Fine grain

Data redundancy

The 'Push' model using the original PIPs has a high rate of data redundancy (Damodaran, 2004). Two such PIPs in the 'Push' model could give rise to as high as some 50% of data redundancy. Since a complete PIP business transaction could consume hundreds of gigabytes, the usage of structured data-blocks in models nos. 2, 3, and 4 listed in Table 1 could reduce the rate of data redundancy to a large extent. Comparatively, model no. 5 has the lowest rate of data redundancy.

In the 'Push' model, the PIP formats are established by RosettaNet in the Trading Partner Agreement (TPA), while the data-blocks in models nos. 2, 3 and 4 are determined between the trading partners and fixed in the TPA. However, in model no. 5 the data fields extractable through the Web services could be determined dynamically during the execution. As such, model no. 5 has the lowest data redundancy. If the users do not wish to deploy the services of an intermediary party, models nos. 3 or 4 could be the wise choices.

Fidelity

The 'Push' model using the original PIPs has a higher rate of data accuracy or higher fidelity, since the sender pushes all data to the receiver. However, in the 'Pull' model, the receiver may not be aware of the new updates. As such, the sender has to notify the receiver on the new updates and request the receiver to pull them. This problem does not arise in the 'Push and Pull' model because the sender always pushes the latest data-blocks to the receiver. In the case of model no. 5 listed in Table 1 where Web services are used, the receiver would be notified whenever there are new updates.

Accessibility

Accessibility is the availability of the receiver's server for the 'Push' model or the sender's server for the 'Pull' model. In other words, both the partner's server(s) must be online. Among the five models, models nos. 4 and 5 listed in Table 1 have the highest accessibility since it is assumed that the intermediary server(s) has multiple levels of server backup and will always be communicating with the trading partners. In these two models, the sender always pushes the updates to the intermediary server which in turn would notify the receiver for some data pulling actions. Models nos. 1 and 2 have low accessibility because the small and medium-sized industries do not have the budget to purchase additional server facilities.

Personalizability

Personalizability is the ability to suit one's needs and wants. Such capabilities are provided through the Web services implemented at the intermediary party's Website

where the partners interact, and through all the available communication channels and media. The advantage of using a third party's services is that a common set of programmable services are stored at the server which provide 24-hour accessibility. At least two aspects of the services could be personalized namely, timing and content. The sender can push data to the server and the receiver can pull the smallest possible units of data from it at the most convenient time.

Models nos. 1 and 2 listed in Table 1 are inflexible in both timing and content due to the predefined PIPs format, while model no. 5 is the most flexible in timing and could provide the smallest possible units of data through the Web services.

Conclusion and Future Work

With the ulterior aim of researching into the personalizable Web services for the smaller industries, this paper analyzed the original 'Push' model in the RosettaNet Partner Interface Processes, and generally compared it with the proposed 'Pull' model and 'Push and Pull' model. Their performances are assessed and evaluated in terms of data redundancy, fidelity, accessibility and personalizability. It is argued that when the 'Pull' and 'Push' models are combined, the resulting model would enhance the overall document interchange performance and flexibility. The outcome of this research would add value to the provision of personalizable Web services which may eventually converge and/or interface with the new RosettaNet implementation framework in the future. With these personalizable Web services, the users among the small and medium-sized industries would then have better-informed choices of document interchange models to suit their individual interaction styles.

Further research on industrial gap analysis is required to possibly standardize some of the private business processes among the smaller non-Electronics industries so that they will be more ready to use the new RosettaNet implementation framework and other open document interchange frameworks or standards.

References

Accenture. (2004). *Breaking away: Creating business value through information technology*. Retrieved October 1, 2007, from http://www.accenture.com/Global/ Research_and_Insights/ Outlook/By_Alphabet/BreakingTechnology.htm

Bhide, M., Deolasee, P., Katkar, A., Panchbudhe, A., Ramamritham, K. & Shenoy, P. (2002). Adaptive push-pull: Disseminating dynamic web data. *IEEE Transactions on Computers*, *51*, 652-668.

Chung, J. Y. (2002). Beyond e-marketplace & next generation e-business: Grid, autonomic computing & web services. *The 4th International Conference on Electronic Commerce, Oct. 24, 2002,* Hongkong.

Damodaran, S. (2004). B2B Integration over the Internet with XML: RosettaNet successes and challenges. *The 13th International World Wide Web Conference on Alternate Track Papers & Posters.* New York, NY, USA, ACM press, 188-195.

Damodaran, S. (2005). RosettaNet: Adoption brings new problems, new solutions. *The XML 2005 Conference.* November, Atlanta, USA, 1-14.

Hanson, W. A. & Kalyanam, K. (2007). *Internet marketing & e-commerce.* Canada: Thomson South-Western.

RosettaNet. (2002). RosettaNet: implementing e-business processes on a global scale. *RosettaNet Executive Insights*.

RosettaNet. (2007a). *PIP directory*. Retrieved September 1, 2007, from http://www. rosettanet.org.my/PIP_Directory.aspx

RosettaNet. (2007b). *RosettaNet technical overview*. Penang, Malaysia: RosettaNet Engineering.

RosettaNet. (2007d). *Why RosettaNet*. Retrieved September 20, 2007, from http:// www.rosettanet.org.my, 2007

Schönberger, A. (2006). Modelling and validating business collaborations: A case study on rosettanet. *Bamberger Beiträge zur Wirtschaftsinformatik und Angewandten Informatik*, *65.*

Surfcontrol. (2007). *Excessive spam e-mails were beginning to slow down the network, jeopardizing fast turnarounds on customer projects*. Retrieved September 1, 2007, from http://www. surfcontrol.com/ general / assets /case_studies / imagemaster.pdf, 2007.

Ting, T. T., Khoo, K. T. & Cheah, Y.-N. (2007). Approaches to enhancing rosettanet standards component. *Computer Science Postgraduate Colloquium*, 8-9.

Wikipedia. (2007a). *Electronic commerce*. Retrieved September 20, 2007, from http://en.wikipedia.org/wiki/E-commerce.

Wikipedia. (2007b). *Electronic data interchange*. Retrieved July 25, 2007, from http://en.wikipedia.org/wiki/Electronic_Data_Interchange.