An XML-based data interchange protocol and supporting systems for online customs declaration

Li, Y.; Wu, X.; Shen, W.; Chai, Y.

Publisher's version / Version de l'éditeur:

2009 IEEE International Conference on Systems, Man, and Cybernetics [Proceedings], pp. 4094-4099, 2009-10-11

An XML-based data interchange protocol and supporting systems for online customs declaration

NRCC-51245

Li, Y.; Wu, X.; Shen, W.; Chai, Y.

October 2009

A version of this document is published in / Une version de ce document se trouve dans:
2009 IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, Texas, October 11-14, 2009, pp. 4094-4099.

An XML-based Data Interchange Protocol and Supporting Systems for Online Customs Declaration

Yinsheng Li[1], Xiaoyan Wu[1], Weiming Shen[2], Yueting Chai[3]
[1]*Software School, Fudan University, Shanghai 200433, China*
[2] *National Research Council of Canada, London, Ontario, Canada*
[3]*National CIMS ERC, Tsinghua University 100084, China*

* Corresponding author. E-mail:

Abstract

Online customs declaration is becoming popular for international logistics through airports and seaports. This paper proposes a dedicated data interchange protocol and presents the supporting systems being developed. Compared with the traditional Electronic Data Interchange, the proposed protocol and systems is based on XML, and supports RFID, Harmonized System codes, and image processing. The protocol and systems have been used by more than 30 logistics companies and airports in China. Among them are top logistics services companies like FedEx, DHL, UPS and TNT.. The protocol is being considered as a national standard for customs declaration in China and has a potential to be accepted internationally based on China's best practices.

Keywords: Online Customs Declaration, RFID, Harmonized System, Electronic Data Interchange, Logistics Services.

1. Introduction

Online customs declaration is becoming popular with international logistics companies for providing cross-boarder services through airports and seaports. There are two parties involved in the declaration: logistics companies such as FedEx, DHL and UPS, and public departments such as inspection and customs authorities.

In China, public departments of inspection, quarantine, and customs have been encouraged to use computers and information systems to streamline the application and inspection processes, improve their processing efficiency and accuracy, and avoid illegal human interventions in the quarantine or customs policies during declarations. The Department of Chinese Inspection and Quarantine has, therefore, developed a distributed information system, i.e., CIQ 2000. CIQ 2000 has been put into use by all ports of China, to facilitate inspection and quarantine processing [1].

Logistics companies such as FedEx, DHL and UPS, have also been equipped with powerful information systems to track their shipments and maintenance and performance records, so that their clients can check on the status and whereabouts of their parcels.

The declaration of customs, quarantine and inspection need the logistics companies to transmit their shipment data to the public departments for approval. To improve the declaration process, it is a prerequisite and a major challenge for the two parties to be interlinked and exchange declaration data via the systems. Three elements are required to make this happen: (i) a declaration system to collect, verify and submit declaration data at the logistics side; (ii) a processing system to receive, process and respond to the declarations at the public department side; (iii) an open data interchange protocol and mechanism to link the two systems and to exchange data in an efficient, secure and complete way.

To design and develop these three elements, we conducted hundreds of interviews with officers from Shanghai International Airport Inspection and Quarantine Bureau, and FedEx, DHL, UPS, and TNT. With their full support and cooperation in the past three years, we proposed and developed a dedicated data interchange protocol and two supporting systems, i.e., a declaration system and a processing system. Compared with the traditional Electronic Data Interchange, the proposed protocol is based on XML and with support of RFID, Harmonized System codes, and image processing. The two supporting systems have been implemented to function as both a desktop applications or Web services-based middleware. At the time of writing this paper, the proposed protocol and middleware have been used by more than 30 logistics companies and airports in China, including top logistics companies like FedEx, DHL, UPS and TNT in their businesses, and Shanghai International Airport. The related methods have patents pending. The proposed protocol is being considered as a national standard for entry & exit declaration in China and has the potential to be accepted internationally, based on China's best practices.

The rest of this paper provides a comprehensive description of the proposed protocol and supporting systems (or middleware). Section 2 gives a survey on the related work and technologies. Section 3 describes declaration-related concepts and processes. Section 4 presents the protocol mechanism. Section 5 presents the basic design of the supporting systems. Section 6 discusses the protocol and system's application and feedback. Section 7 concludes the work.

2. Motivation and Related Techniques

We conducted interviews with a good number of logistics companies and public departments to investigate the existing declaration systems and processing systems, including Shanghai International Airport Inspection and Quarantine Bureau, and Fedex, DHL, UPS, and TNT at their bases in China. The findings were not encouraging. Most of the logistics companies have been using FTP clients to submit declaration data. Some companies collect data manually from their tracking systems. Other companies can produce declaration data using their management systems. As for public departments, a few of them applied information systems to process. They moved data from the FTP sites and put them into their processing systems to get decision-making support. Very few of them are developing in-house pre-processing systems to classify declarations and facilitate the approval process.

Based on the above observation of the existing declaration and processing systems, we propose to develop two functional systems: a declaration system for logistics companies, and a processing system for China's public departments of customs, inspection, and quarantine. A data interchange protocol is also required for data exchange between the two systems. The protocol should support emerging RFID labels, image checking, and Harmonized System codes. The RFID and Harmonized Systems have been accepted to be applied in customs declaration. Chinese authorities have required logistics companies to provide scanned parcel images for possible inspection or auditing. The traditional EDI does not support the three features for customs declaration though the features have been put into force by national policies. .

The main functions and services of the proposed systems are implemented and published as Web services. Declaration and processing Web services can be implemented as middleware for the declaration and processing systems to be integrated into the information systems of logistics companies and public departments. The emerging technologies of Web services and the Semantic Web present a promising solution to the rapid integration of heterogeneous applications, and promote the adoption of a service-oriented architecture in the e-business systems. They are expected to transform a group of loosely coupled Web business applications into a collaborative distributed business network [9].

The related technologies and work related to the proposed protocol is EDI, XML, RosettaNet, cXML and ebXML. All of them are e-business frameworks and allow organizations in the e-business chain to interact with each other [8]. However, customs declaration is a special interaction process governed by public departments rather than an open business process. The RosettaNet, cXML and ebXML, which allow for open e-commerce are not suitable. The declaration data interchange protocol needs basic data formatting and a transporting technology. The most popular data interchange technology used by entry & exit departments is EDI. EDI is a major B2B data interchange standard in past decades. EDI has many successful cases. It is well-accepted among transnational companies. However, high price and private standards are drawbacks of EDI. For letting the information systems of business partners accomplish electronic business communication, semantic inter-operability is necessary to ensure that exchange of information makes sense – that the provider and the receiver of information have a common understanding of the "meaning" of the requested services and data. Traditional EDI is not sufficient to solve electronic business communication problems in an open and dynamic environment [4].

Extensible Markup Language (XML) provides an affordable and flexible solution. With XML, data can be exchanged between incompatible systems. In the real world, computer systems and databases contain data in incompatible formats. One of the most time-consuming challenges for developers has been to exchange data between such systems over the Internet. Converting the data to XML can greatly reduce this complexity and create data that can be read by many different types of applications [5] [6].

The proposed declaration data interchange protocol will use XML to have the capability of semantic and flexible integration. Considering EDI has been used for many years, the proposed data interchange protocol uses the basic EDI concepts, based on XML grammar, to make the proposed protocol more professional and easily accepted. The protocol, therefore, is an XML-based EDI for this special application. As a result, it is convenient to transforms the EDI data to XML data, or vice-versus.

The Harmonized System (HS) is a systematic and multi-purposed classification directory for international trade [http://www.wcoomd.org]. The HS serves as a universally accepted classification system for goods so that countries can administer customs programs and collect trade data on exports and imports. It was designed to replace the local systems used by countries allowing them to have a common classification system by which to track trades and apply tariffs. The HS has been adopted by the majority of the world's trading countries as being the basis for entry and exit monitoring and trade statistics. Each product can be assigned with a unique identification number, which are usually 6-10 digits. The basic HS uses a 6-digit number to identify basic commodities. Each country is allowed to add additional digits for statistical purposes. For example, Canada uses an additional 2 digits for exports and an additional 4 digits for imports. The USA uses a 10-digit system for both exports and imports. China use 4-14 digits for both exports and imports.

Commodities (parcels) can be applied with different entry and exit policies based on their unique HS codes. The HS codes are the key to making the declaration process automated, if only logistics companies could provide HS codes associated with their parcels. However, in many cases, logistic companies do not provide HS codes when reporting to customs or quarantine officials. Customs and quarantine departments then have the tedious job of identifying the relevant codes for the relevant products. This makes the inspection process complicated and inefficient. Current inspection or customs information systems find it difficult to automatically and intelligently inspect or check products without HS codes. Moreover, information systems are supposed to be updated with dynamic inspection, quarantine and customs policies, and these newly released policies are based on HS codes. Since these information systems are unable to fully exploit HS codes, they are inflexible and difficult to upgrade and maintain. In the majority of cases, human experts are required to file entries for HS codes. However, for those individuals with little experience of

customs inspection, providing the correct HS codes can be extremely difficult and time consuming. Logistics companies have the same problem because export goods are sometimes beyond their knowledge to classify.

To utilize HS codes and make the declaration process automatic and efficient, there are 2 prerequisites, i.e., there should be an automatic HS codes retrieving system to help public departments or logistics companies to produce HS codes automatically, and the data interchange protocol should support the HS codes. In our previous work, a Web service-based HS coding service using machine learning and semantics mining has been developed, patented and used by some Chinese public departments and companies like FedEx [2]. In this work, the HS codes will be incorporated into the proposed data interchange protocol.

RFID (Radio Frequency Identification) is a technology similar in theory to bar code identification. One of the differences between RFID and bar code technology is that RFID eliminates the need for line-of-sight reading of the bar coding system. RFID scanning can be done at greater distances than bar code scanning. RFID has rich serial number resources to ensure every single item to has a unique ID. With a registration database storing commodity information associated with each commodity ID, every commodity item can have a complete description to be retrieved and shared between business partners or their clients [3].

One of the RFID's major applications is in the logistics industry. Logistics companies can track the shipment of their parcels once they put RFID tags on the parcels and place RFID readers at the gates where the tagged parcels go through. It is valuable to use RFID during the declaration process. Besides the tracking function, there are several features to be accomplished by RFID. RFID is a prerequisite for inspection and customs authorities to check the details of the posted parcels. Before using RFID, officers need to use barcode readers to touch parcels to get their barcode numbers. It is inconvenient and inefficient, either for warehouse management and maintenance, or for parcel details checking. What makes a big difference is that RFID allows officers to be able to get information for each parcel when they pass through the transmission straps at a high speed. Furthermore, RFID tags are associated with rich commodity information so that the processing system can use this information to get more precise HS codes (The HS coding service can produce more precise codes when more detailed product description is provided). Furthermore, the RFID tags have the potential to replace the traditional parcel IDs for commodity declaration, and make the IDs recognized internationally.

3. Concepts and Processes for Online Customs Declaration

Transnational logistics is one of the key characteristics of modern logistics. As mentioned above, the entry & exit declaration process is a critical node for transnational logistics. The declaration processes are usually composed of interactions between logistics companies and public departments for inspection, quarantine or customs. The declaration can take place at airports, seaports, or railway stations. Among them, express declaration between logistics express companies and public departments of inspection and quarantine is very common, the most complicated, and provides a good example for the purpose of identifying the requirements and illustrating the concepts.

The concepts involved in express declaration include:

i) Warehouse receipts. Produced and provided by logistics express companies. On the header of a warehouse receipt are messages for the express note by a logistics express company. An express note can have more than one subordinate note, and each subordinate note ID is associated with a single note ID. The body of a warehouse receipt can hold rough messages for a subordinate note. When using XML to construct the note, an XML file of a warehouse receipt includes one warehouse receipt header and more than one warehouse receipt body area.

ii) Declaration receipts. Produced and provided by logistics express companies. A declaration receipt holds the complete details of one or more subordinate notes. Using XML to construct the note, an XML file of subordinate note includes one declaration receipt header and zero or more declaration receipt body area. The declaration receipt header holds the statistics and image of all commodities with this declaration receipt.

iii) Transition receipts. Produced by public departments for transition between inspection and customs. A transition receipt holds the data including transportation ID number, carrier type, shipment mode, entry or exit location, company number, export/import type, total weight, carrier, and a count of the declaration receipts with express note.

The declaration processes are governed by public departments. According to the public departments, there are two classes of declaration processes: customs declaration process and inspection processes. And there are two types of inspection processes: fast-track inspection and routine inspection. To sum up, there are 3 levels of declarations, which are applied with different levels of declaration sequences and data requirements, as illustrated in Figure 1.

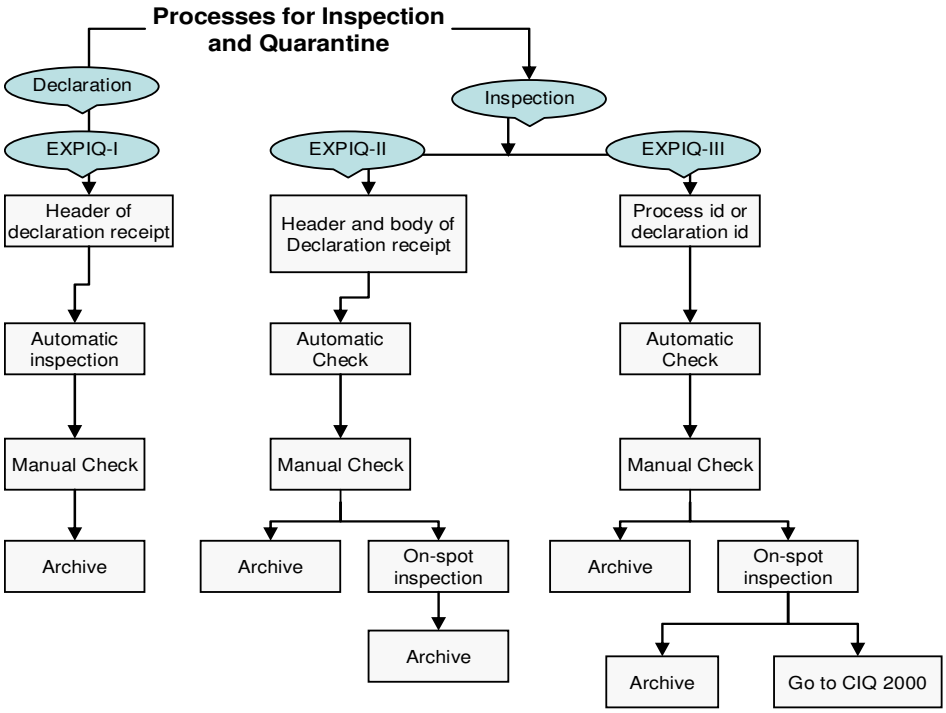

Figure 1. The 3 levels of express declaration processes

i) The first level of declaration, named as EXPIQ-I by Chinese public departments, is for those goods that do not need inspection and quarantine. They are required to only provide headers of declaration receipts.

ii) The second level of declaration, named as EXPIQ-II, is for those goods which are samples, gifts, private commodities, not-for-sale commodities, and the commodities with no more than RMB 2000 of value. They are required to provide headers and bodies of declaration receipts for inspection.

iii) The third level of declaration, named as EXPIQ-III, is for those goods which do not fall into EXPIQ-I or EXPIQ-II. They do not need to provide declaration data, and can go to the CIQ 2000 of China for declaration, as mentioned at the beginning of this paper.

4. Proposed Data Interchange Protocol

4.1 Brief description

The uniqueness of the proposed protocol is its integration with the Harmonized System, RFID, and image processing.

The proposed data interchange protocol carries RFID data both for the items and the parcels. This will help the public departments get more information and conduct various checks and monitoring. A pilot project has been completed to use RFID to demonstrate the above functionalities and features, with the support of Shanghai International Airport Inspection and Quarantine Bureau, FedEx, and DHL.

In a proclamation by Chinse public departments, logistics are required to provide scanned images of their express parcels to be checked in case the information provided does not satisfy declaration approval. To this end, the proposed data interchange protocol should carry image data of the express packages.

The functional requirements, processes and technical features of entry & exit declaration have been investigated, optimized and concluded, based on hundreds of interviews with the officers from Shanghai International Airport Inspection and Quarantine Bureau, FedEx, DHL, UPS, and TNT. This section provides summarized design details of the proposed data interchange protocol and the two supporting systems for commodity declaration, i.e., a declaration system and a processing system. The express declaration system is used as an example for illustration.

As illustrated in Figure 2, the declaration system is operated by logistics users and produces XML-based warehouse, declaration, and transition receipts. The declaration system also processes scanned images and transforms them into binary codes. The XML messages are sent out via FTP on aWLAN (Wireless Local Area Network). The processing system is a part of the office information system at the public departments. The processing system receives the receipts with declaration information, including HS codes, RFID numbers, and images in an XML file. The receipts are preprocessed to restore the images, check commodity information, using RFID numbers, and check policies, using the provided HS codes. There are special services to process images, HS codes, and RFID numbers. The return receipts are

stored at the secure repositories and retrieved by the declaration system. To sum up, the declaration system operates asynchronically with the processing system. The declaration system pushes the declaration receipts in and retrieves them from the processing system. It is advisable that the communication protocol is FTP on the proprietary wireless local area network, for security reasons. The FTP is equipped with accounts and passwords for logistics personnel. Moreover, both the declaration system and the processing system have published their main functions as Web services. Logistics companies and public departments can develop Web services clients to get their own systems, or integrate them into their office information systems.

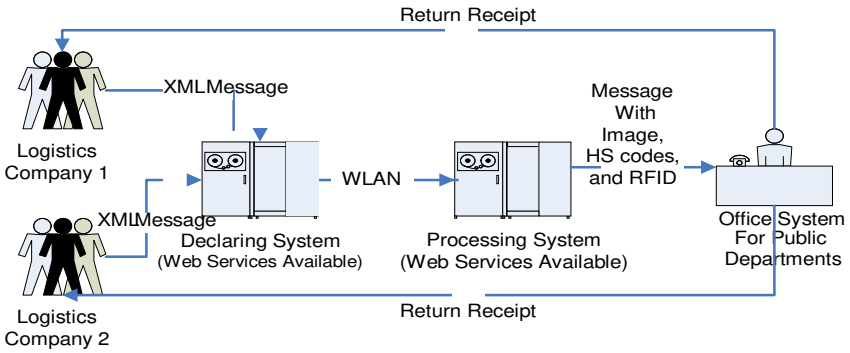

Figure 2. The communication architecture of the express declaration with the protocol

4.2 Concepts and mechanisms

1) XML files to be submitted by logistics companies:

There are three types of data files submitted by logistics companies: warehouse receipts for archival, declaration receipts for declaration, and transition receipts for declaration. Among them, warehouse and declaration receipts have both header and body data, and transition receipts only have header data. The header and body data of warehouse and declaration receipts can be put into a single XML file and submitted.

2) Return receipt to be received by logistics companies:

The processing system at public departments processes the submitted warehouse receipts for archival, declaration receipts for declaration, and transition receipts for declaration. The corresponding return receipts for warehouse and inspection are output and moved into the predefined directory. The declaration system at logistics companies gets them back using the FTP protocol.

3) XML Schema based messages, data types and usages:

There is a long specification for definitions of messages, data types and their usages involved in the data interchange during express declaration. Table 1 and Table 2 are part of the protocol definition related to image, RFID, and HS processing. For example, the defined image will be transformed from various image formats (jpg and tif are supported at this time) into binary codes of base64. The binary codes will be incorporated into an XML file, and would be carried to the public departments. With the decoding functions of base64, the binary codes can be restored. Moreover, the image can be optional for being carried by an XML message, or transmitted separately. For separate transmission, the image is required to be associated with its declaration receipt to be identified at the other end.

Table 1. Image definition attached in the Header of the declaration receipt

Fields	Data Type	Scope	Annotation
Image_Type	Enum, “0”, “1”, “2”, “3”	Not empty	“0” for no image, “1” for image embedded in this file, “2” for image will come on request, “3” for image at logistics company and public departments need go for it.
Image_URL	String(1-60)		Location when the image stored. Required when Image_Type is “3”.
Image_Content	Binary Dataflow		Image name, type and body. Required when Image_Type is “1”. Fmt: image format (jpg, tif, expandable)

Table 2. Body of the declaration receipt (in the uploaded file)

Fields	Data Type	Scope	Annotation
Message_Type	Enum, “N”, “A”, “M”, “D”	Not empty	“N” = new added, “A” = added, M = modify, “D” = delete

Voyage_NO	String(1-32)	Not empty	Carrier No.
Bill_NO	String(1-32)	Not empty	
Goods_Name	ComplexType, Goods_Name	Not empty	Goods_Name_EN; Goods_Name_CN
Goods_HS	String(10)	Not empty	HS codes with 10 digits
…	…	…	…
RFID	Enum, “EPC”、“UID”		The used RFID specification
RFID_NO	String(48-256)		RFID No.,
Reserved	String(1-100)		3 reserved fields in total

5. A Declaration System Based on the Protocol

5.1 Structure of a declaration System

Based on the proposed protocol, a declaration system has been designed to process the expresses, shipment and declaration information to produce XML-based receipts on the predefined protocol schema. It is also required to upload, retrieve and handle the return receipts and error messages. As shown in Figure 3, there are five functional modules in addition to home interface, login module, and configuration module.

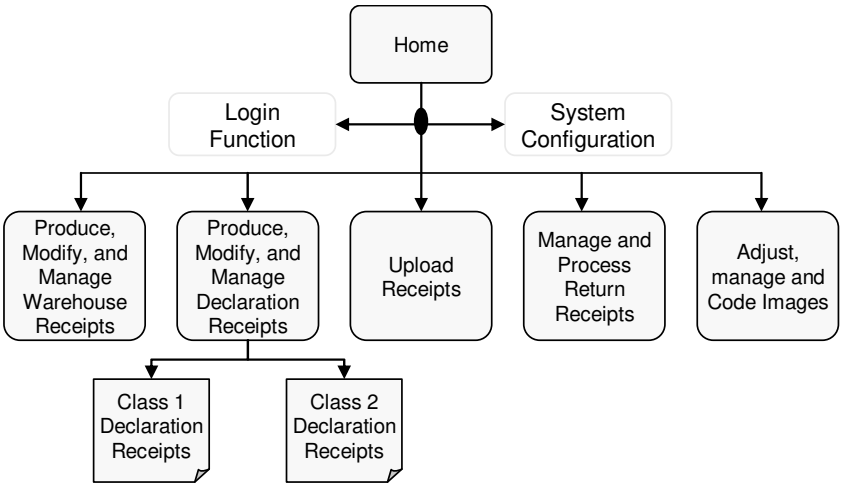

Figure 3. Modules of the declaring system

The module for warehouse receipts translates an EDI file (with extension of .MFT) by airport customs departments into an XML file, based on the proposed protocol schema. The module provides a user interface to navigate, modify and submit the produced XML file.

The module for declaration receipts translates an EDI file (with extension of .ENT) by airport customs departments into an XML file based on the proposed protocol schema. The module can attach image information if the type of declaration receipts is EXPIQ-I or EXPIQ-III. The module also invokes the mentioned Web service-based HS coding service [2] to get HS codes for the commodities.

The module for return receipts retrieves the return receipts and shows them on the user interface. There are two options to display the return receipts. Users may input the express note and subordinate note to get the return receipts. They may also get the complete or part of the return receipt list based on the date or categories.

The module for message upload is embedded in the modules of warehouse receipts, declaration receipts and image processing. The three modules invoke it to upload the produced XML messages to the processing system. The upload module uses the FTP protocol to upload and has a user interface like popular FTP clients,. The network is of a proprietary wireless local area network. The proposed protocol schema is used to validate the uploaded XML files.

The module for image processing is used to transform various image formats (jpg and tif are supported at this time) into binary codes of base64. The source images are produced by logistics companies using scanners or digital cameras. The binary codes will be incorporated into an XML file, and will be uploaded to the public

departments through the upload module. A translation function has been developed to process the jpg and tif images into binary codes. Figure 4 is an example from a tiff file to binary codes in the XML file.

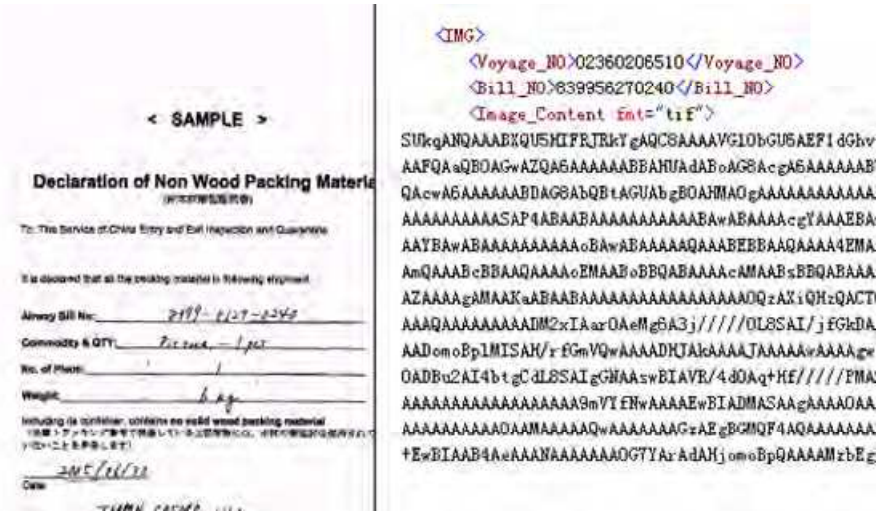

Figure 4. From a .tiff image into binary flow in XML

5.2 Functional design of the processing system

The processing system is a counterpart of the declaration system. It is part of the office information system of public departments. The processing system needs to scan the designated communication port and receive the receipts with declaration information. It invokes the related services to process and produce return receipts. The return receipts are stored at the tree-like directories behind a firewall.

Other specialized services are running within the office information system, e.g., HS-based policy checking service, RFID-based sample checking service, image navigating service, image decoding service, and monitoring service. The processing system invokes the HS-based policy checking service to determine the applied policies, the RFID-based sample checking service to inspect if there are prohibited commodities in the parcel (each single commodity in the parcel may be checked even it does not have a RFID tag), image navigating service to help check the precise parcel information, and image decoding service to restore the received binary image into the image at their source format.

The declaration and processing systems have been developed in a J2EE environment, i.e., Java and Eclipse. The database and data operation are based on PL, XML, Java, JDBC and Oracle. The resulting system was debugged and tested on an IBM X-Series server.

Among the clients, DHL China put the protocol and declaration systems into use in October 2006. Figure 4 shows snapshots from the declaration system of DHL China. Shanghai International Airport Inspection and Quarantine Bureau started the processing system in September 2006.

Feedback from DHL China, FedEx China and Shanghai International Airport is encouraging. The user reports show that the declaration process has been speeded up by 200% and the declaration cost has been dramatically reduced. Take Shanghai International Airport Inspection and Quarantine Bureau as an example. Their report shows that they have achieved mass processing and automatic processing with the processing system and special services. Human intervention has been reduced. Moreover, one of their departments had saved 5 million RMB of processing cost for the year, as can be seen in Table 3 from their report.

6. Evaluation and Feedback

Table 3. Profits and cost reduction by the processing system and services

Profit Type	Human Resources	Social Benefits	Economic Benefits
Amount	5 million RMB for the year. Will beyond 10 million once used in two departments.	Declaring process speeds up by three times	A million RMB/Year
Description	The system has been used at the Department of Express, and released 20	Express companies also speed up declaration process by three	Paper, transportation, and office savings by online

	processing officers which cost 5 million RMB.	times	processing

7. Conclusions

Based on hundreds of interviews with the officers from logistics companies and related public departments in China, the functional requirements, processes and technical features of entry & exit declaration have been investigated and concluded. A special data interchange protocol and two supporting systems, i.e., a declaration system and a processing system, have been proposed and developed. Compared with the traditional Electronic Data Interchange, the proposed protocol is based on XML and with support of RFID, Harmonized System codes, and image processing. The proposed protocol and supporting systems have been used by more than 30 logistics companies and airports in China.

8. Acknowledgement

We would express our grateful thanks to the officers from Shanghai International Airport Inspection and Quarantine Bureau, Fedex, DHL, UPS, and TNT. It is their full support with business knowledge, funding, and cooperation that has made this work a success. Special thanks are given to Mr. Weizu Zhu, Mr. Chao Wang, Mr. Yue Zhang, and Mr. Dongqi Chen.

References

[1] Binyu Zang, Yinsheng Li, Wei Xie, Zhuangjian Chen, Chen-Fang Tsai, Christopher Laing. An ontological engineering approach for automating inspection and quarantine at airports. Journal of Computer and System Sciences. (Article in Press).

[2] Yinsheng Li, Zhanxin Ma, Wei Xie, Chris Laing. Inspection-oriented coding service based on machine learning and semantics mining. Expert Systems with Applications. Volume 31, Issue 4, November 2006, pp. 835-848.

[3] Yi Zhi Zhao, Oon Peen Gan. Distributed Design of RFID Network for Large-Scale RFID Deployment[J]. Industrial Informatics, 2006.08.

[4] Wilhelm Hasselbring, Hans Weigand. Languages for electronic business communication: state of the art. Industrial Management & Data Systems. Volume: 101 Issue: 5. 2001 pp.217 – 227.

[5] David A. Koonce and Suporn Chenhansa. An Approach In Applying XML To Logistics Communication. http://citeseer.ist.psu.edu/540268.html.

[6] XML/EDI Group [M]. http://www.xmledi-group.org.

[7] Yinsheng Li, Xiaohua Lu, Kuo-Ming Chao, Ying Huang, Muhammad Younas. The realization of service-oriented e-Marketplaces. INFORMATION SYSTEMS FRONTIERS. 8 (4). SEP 2006. pp.307-319

[8] Medjahed, B. Benatallah, A. Bouguettaya, A. H. H. Ngu, and A. K. lmagarmid. Business-to-business interactions: issues and enabling technologies. VLDB Journal, 12(1):59–85, 2003.