

XML Schema Design and Management for e-Government Data Interoperability

Thomas Lee¹, C.T. Hon² and David Cheung¹

¹University of Hong Kong, China

²Macau University of Science and Technology, China

ytlee@cecid.hku.hk

cthon@must.edu.mo

dcheung@cecid.hku.hk

Abstract: One-stop public services and single window systems are primary goals of many e-government initiatives. How to facilitate the technical and data interoperability among the systems in different government agencies is a key of meeting these goals. While many software standards, such as Web Services and ebXML, have been formulated to address the interoperability between different technical platforms, the data interoperability problem remains to be a big challenge. The data interoperability concerns how different parties agree on what information to exchange, and the definition and representation of such information. To address this problem, the Hong Kong government has released the XML Schema Design and Management Guide as well as the Registry of Data Standards under its e-Government Interoperability Framework initiative. This paper introduces how the data modelling methodology provided by the Guide can be used to develop data interfaces and standards for e-government systems. We also discuss how the Macao government has formulated their data interoperability policy and has applied the Guide in their situation.

Keywords: e-government data interoperability, XML schema

1. Introduction

Hong Kong and Macao are two special administrative regions (SARs) of China. Under China's one-country-two-systems framework, the two cities are administered by their own governments. The two SAR governments are actively implementing one-stop electronic services for citizens. This way, citizens can find and access public services in a single government portal through the Internet without knowing which government departments offer what services. However, the information systems of different government departments are generally independent and heterogeneous, and one single information system for all departments is unlikely to take place for various reasons. This leads to the requirement to establish an interoperability framework for system integration among different departments. This paper discusses how Hong Kong and Macao SAR governments formulate their interoperability frameworks, particularly in enabling data interoperability among different governments using the Extensible Markup Language (XML) (W3C 2004).

1.1 Hong Kong digital 21 strategy

First launched by the Hong Kong Special Administrative Region Government (HKSARG) in 1998, the *Digital 21 Strategy* (HKSARG 2008) is a strategy paper setting out the blueprint for development of information and communications technology (ICT) in Hong Kong. The paper is maintained by the Digital 21 Strategy Advisory Committee, represented by ICT experts from the government, the industry and the academia in Hong Kong. It is updated regularly to reflect the current technological advancements and changing society needs. The current issue is 2008 Digital 21 published in December 2007; one of its action items is:

"Based on the experience gained through the development of the Interoperability Framework for e-government systems, the Government will collaborate with different sectors to develop industry-specific data standards taking into account international standards and best practices. In the process, we will also attempt to enhance interoperability between the data standards of different sectors to facilitate the provision of joined-up, value-added services."

1.2 HKSARG Interoperability framework for e-government

A key business objective of e-government initiatives is to provide client-centric joined-up government services to the public. In HKSARG, public services are offered by different HKSARG Bureaux and Departments (B/Ds) in a distributed fashion. To provide one-stop comprehensive services, seamless

flow of information across individual B/Ds is a prerequisite. In 2003, the HKSARG Information Technology Services Department, now known as Office of Government Chief Information Officer (OGCIO), established the Interoperability Framework (IF) for E-Government (OGCIO 2008) to meet this objective.

IF maintains a collection of technical and data specifications to support different B/Ds to exchange data between their systems in order for implementation joined-up public services. Adopting XML as a key technology to define data exchange interface, IF enables computer systems developed on heterogeneous hardware and software platforms by different B/Ds to interoperate with each other.

A special organization structure called Interoperability Framework Co-ordination Group (IFCG) is formed, with experts from the government and industry, to oversee the IF development. IFCG coordinates with the specialist groups across B/Ds who develop specific interoperability standards for their respective business areas. Being a long-term and on-going strategic effort, IFCG regularly reviews IF every 6-12 months in order to accommodate new business requirements and current technology developments. There is a compliance policy which requires all B/Ds to adopt the IF when implementing new systems that need to exchange data with systems of other B/Ds and external parties. Moreover, IT suppliers and system integrators are also obliged to ensure their solutions provided to the government comply with the IF requirements.

1.3 Three dimensions of interoperability

IF addresses three dimensions of interoperability, namely *technical interoperability*, *data interoperability*, and *process interoperability*. The *XML Schema Design and Management Guide* (OGCIO 2006) describes these terms as follows:

- *Technical interoperability*: agreement on what communication protocol and message format to be used when one party sends information to another; e.g. the purchase order shall be encoded in XML, as defined by a specific XML schema, and XML Encryption and XML Signature shall be applied on certain content components, and the XML message shall be sent via HTTP.
- *Data interoperability*: agreement on what information has to be transmitted from one party to another, and the definition and representation of such information; e.g., the "delivery date" has to be specified on a purchase order, and the definition of "delivery date" is the date on which goods shall be received by the buyer, and the representation of "delivery date" adopts the ISO 8601 standard.
- *Process interoperability*: agreement on how the business activities of the concerned parties affect each other, i.e., the business rules; e.g., when the seller receives a purchase order from the buyer, the seller should accept or reject this order within a specified time period.

To address these dimensions of interoperability, IF maintains two sets of documents. The first set of documents, which include the *HKSARG Interoperability Framework* and *Analysis Underpinning the HKSARG IF Recommendations*, recommend a collection of technical specifications for implementation for application in different interoperability areas. For example, in the Application Integration Domain, Web Services standards SOAP v1.1, WSDL v1.1, and UDDI v2 are recommended for simple functional integration in an open environment while ebXML Message Service (ebMS) v2 is recommended for reliable message exchange between application systems in an open environment for business document oriented collaboration. This means when a new e-government project is recognized to fit a particular interoperability area, the B/D responsible for the project and their IT contractors are required to adopt the recommended technical standards in system implementation.

The other set of documents are collectively called *XML Schema Design and Management Guide* (OGCIO 2006) (Schema Guide). They are developed to address data and process interoperability. In this paper we focus to discuss about the contents of this Guide, which is covered in Sections 2. HKSARG has also created an online portal called the *Registry of Data Standards* to manage the e-government data standards called *Common Schemas*, which is discussed in Section 3. Section 4 lists some e-government projects that have adopted this framework. Finally, Section 6 concludes this paper.

2. XML schema design and management guide

The Center for E-Commerce Infrastructure Development (CECID) of The University of Hong Kong was commissioned to develop the XML Schema Design and Management Guide to help B/Ds and IT

contractors define and adopt e-government data standards. This Guide provides the following to address the data and business interoperability of government-to-government (G2G) and government-to-business (G2B) joined-up services:

- a methodology for business analysts to specify the definitions and representations of information in a consistent and structured way as reusable information models,
- an approach for programmers to convert the information models of the data elements into W3C XML Schema Definition (XSD) (W3C 2004) code,
- the guidelines for the concerted alignment of the definitions and representations of data elements that have potential for reuse in joined-up services, thereby standardizing the XML schema code for these data elements, and
- the guidelines for project teams to adopt suitable concertedly aligned data elements and their standardized XSD code and also to contribute reusable data elements for concerted alignment.

The Guide consists for four parts:

- *Part I: Overview* states the objectives of the Guide and outlines its contents, and describes the data interoperability problems and strategy in e-government development. This part also covers the data interoperability measures and the guiding principles.
- *Part II: XML Schema Design Guide* provides a systematic design methodology for business analysts to model the business process and information requirements in a joined-up project. It also provides the rules for programmers to convert the information models into XSD code.
- *Part III: XML Schema Management Guide* serves as a handbook for various parties to align data elements from different projects to create data standards among B/Ds. It facilitates the development and management of reusable XML Schemas. It also helps project teams understand the process of data alignment, and their role in contributing reusable data elements for concerted alignment.
- *Part IV: Appendices* provide supplementary information to help the readers understand this Guide. Most importantly, there is a case study on using the Guide to define data standards for application for import and export licences for pharmaceutical products and medicines. This part also provides a collection of worksheets for specification of process and information models.

2.1 XML schema design guide

The XML Schema Design Guide (Design Guide) documents a comprehensive XML schema design methodology. There are five core sections: (1) Schema Design Process, (2) Business Process Modelling Methodology, (3) Business Information Modelling Methodology, and (4) XML Schema Definition Development. Unlike other methodologies, such as UN/CEFACT Modelling Methodology (UMM) (UN/CEFACT 2001), the Design Guide provides analysts and developers with an end-to-end and seamlessly integrated schema design methodology. It covers modelling techniques from business process modelling, data modelling to XSD coding.

2.1.1 XML schema design process

The schema design process of the Design Guide is shown in Figure 1. It guides a project team to develop the XML data interface, i.e., XSD, for a government system to exchange data with other government or business systems. The process involves two types of schemas: *Project Schemas* and *Common Schemas*. A Project Schemas only needs to satisfy the specific requirements of a particular project while Common Schemas are reusable across projects and among different B/Ds. Common Schemas are centrally standardized in the government and are shared publicly in the Central Registry. Another difference between Project Schemas and Common Schemas is as follows. A Project Schema is usually document-oriented, like purchase order, because it is ready for use for data exchange. A Common Schema is data-element-oriented, such as address, person name, while it is used as a building block to construct Project Schemas.

The design process can be summarized as follows. First of all, the project team collect and analyze the user requirements to see if there is a suitable industry standard, e.g., Universal Business Language (OASIS 2003). The team start to design a custom schema only when no industry standard is suitable. Then, the business analyst models the business process in which the system exchanges data with other systems, and identifies the business documents involved in the exchange. For each

identified document, the business analyst builds its information model, i.e., the document structure as well as data constraints. The data models are converted into XSD code of the Project Schema by software tools. If the project team anticipate some data elements in the Project Schema are generic enough for reuse by other projects in future, they can propose these data elements for central standardization as Common Schemas.

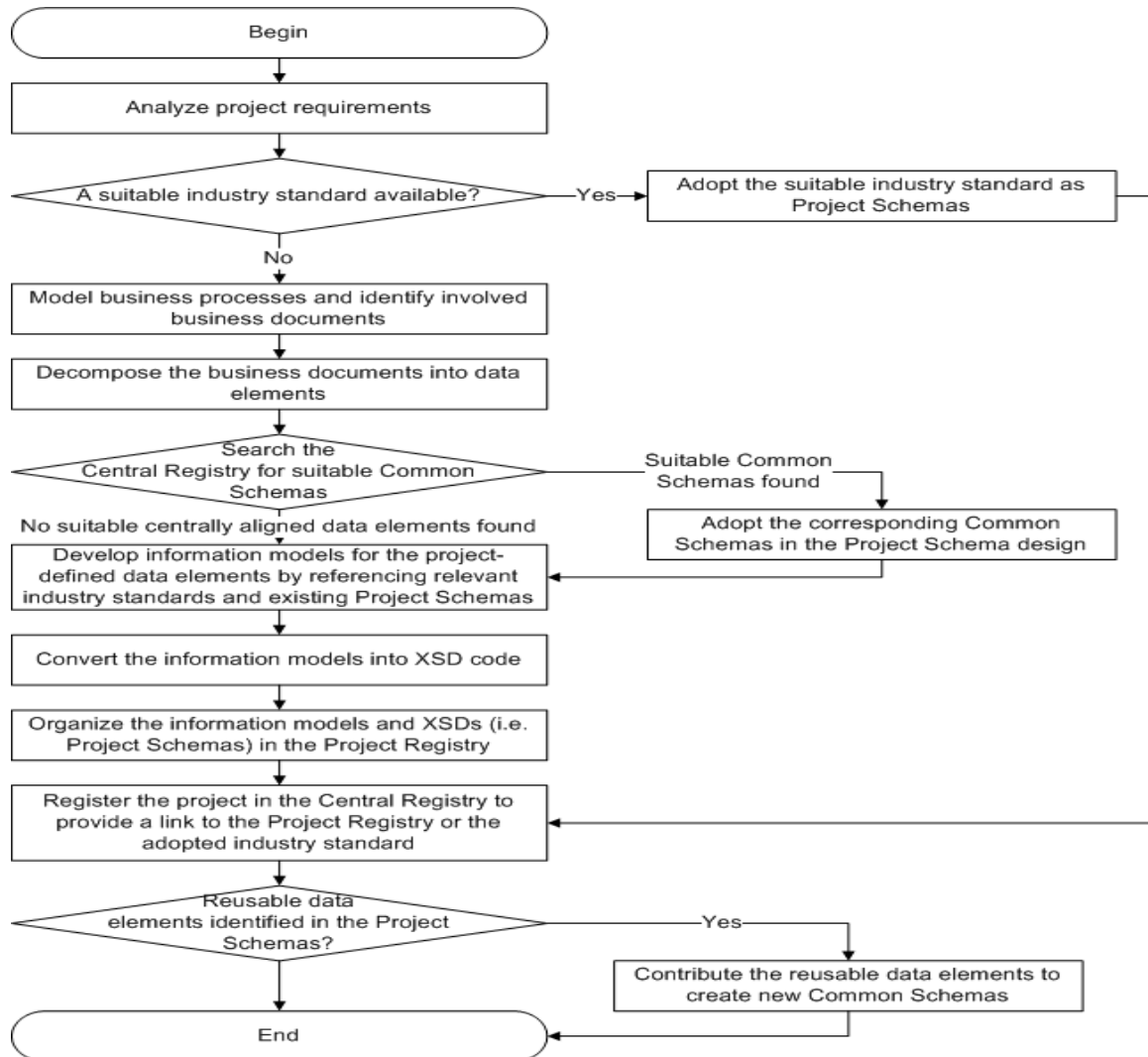


Figure 1: XML schema design process

2.1.2 Business process modelling

The business process modelling (BPM) methodology is derived from the *ebXML Business Process modelling* approach (OASIS 2001). A business process is modelled as a collection of *business collaborations*. A business collaboration is a choreography of business transactions. A business transaction is an abstraction of a single document exchange, which can be a one-way or two-way document flow between two parties involved in the collaboration. When the flows of documents between different parties are identified, these documents will be modelled in the next stage.

The Design Guide provides a set of modelling worksheets for business analysts to fill in the requirements of business collaborations and business transactions. For example, the worksheet in Figure 2 documents an *order entry* business collaboration, which involves a *buyer* as the requesting role and a *seller* and the responding role. Two business transactions, namely *Request Quote* and *Create Order*, are identified in the collaboration. The *Request Quote* transaction involves an exchange of two documents *Quote Request* and *Price Quote* while the *Create Order* transaction is an exchange of *Purchase Order* and *Order Confirmation* documents. These documents will be used for business information modelling. The process model can be specified as a UML activity diagram.

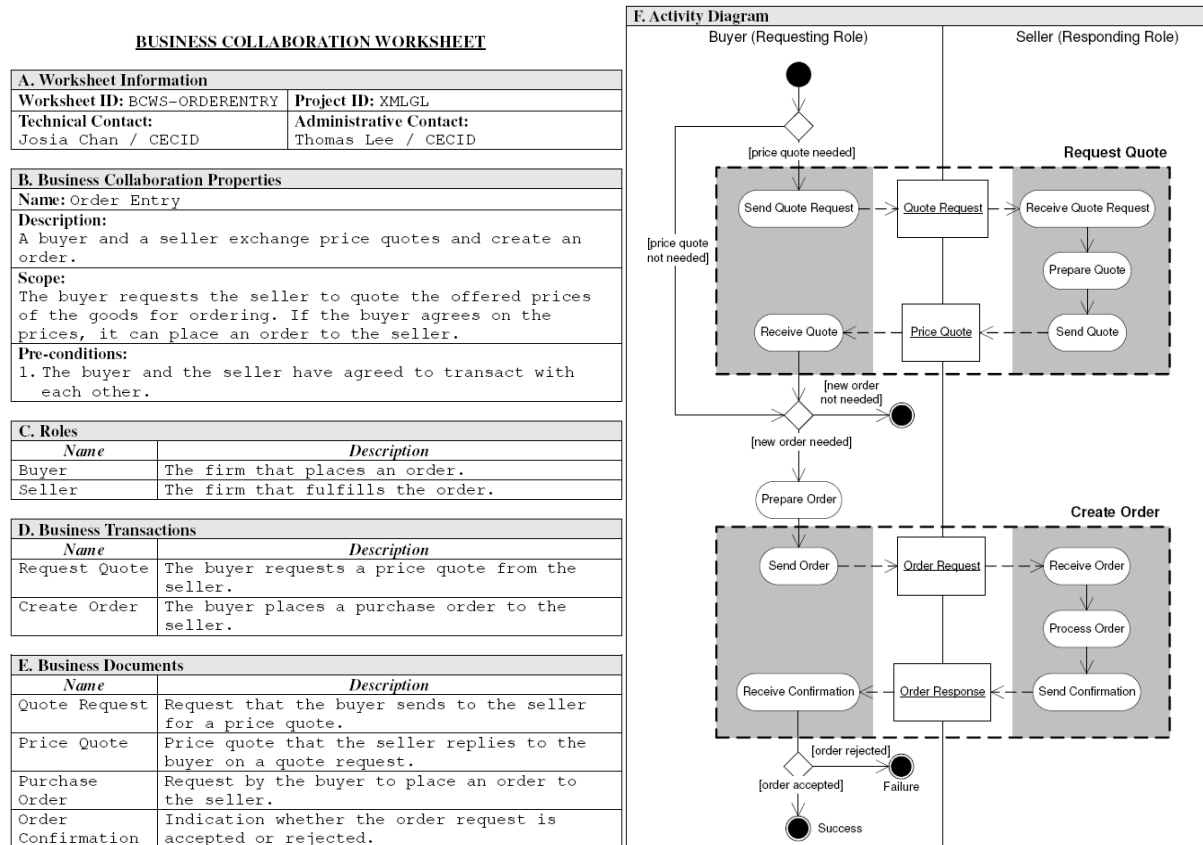


Figure 2: Sample modelling worksheet for a business collaboration.

2.1.3 Business information modelling

The business information modelling (BIM) process gathers the requirements for the business documents identified in the business process modelling. These requirements are usually in form of paper and electronic copies of those documents that are currently exchanged manually or in old electronic fashion. The modelling methodology simplifies and extends the *ebXML Core Components Technical Specification (CTS)* (UN/CEFACT 2003) and allows business analysts to specify documents and data elements as the following information models:

- *Business Document*, modelling an electronic document as a unit for business information exchange; a root Aggregate Business Information Entity is used to provide the representation of the document,
- *Aggregate Business Information Entity (ABIE)*, modelling an object class and aggregates Basic and Association Business Information Entities as the properties.
- *Association Business Information Entity (ASBIE)*, modelling a complex property in an object class.
- *Basic Business Information Entity (BBIE)*, modelling a singular property in an object class.
- *Core Component Type (CCT)*, modelling a basic business data type as a building block for the above models.

BIM provides a set of Core Component Types (CCTs) as the basic business data types for building other the information models. In addition to the CCTs proposed in CTS, e.g., *Amount*, *Date Time*, *Code*, *Quantity*, *Measure*, etc., BIM provides new data types, e.g., *Count* (integer *Quality*), *Boolean* (true or false), and *External*. The *External* type allows the data structure to be defined outside the pre-defined set of CCTs by an external XSD complex type. The relationships of these information models can be represented by a UML class diagram like Figure 3.

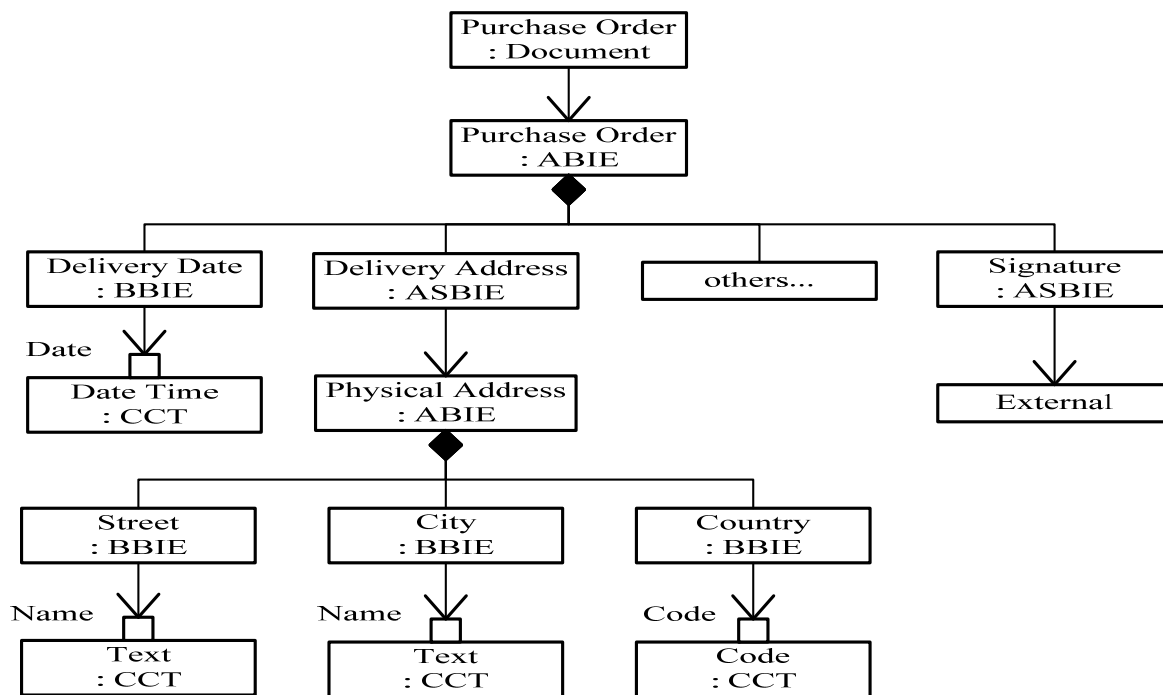


Figure 3: Sample information model for purchase order documents

common_schema																			
A		B		C		D		E		F		G		H		O		P	
Insert Data Item...		Select Data Item...		Show / Hide Less Commonly Used		Export to XML...		Import from											
Dictionary Index						Dictionary Information				Object Class and Properties									
UID	Dictionary Entry Name	Ver.	Maturity Level	Business Terms	Item Type	Item Definition		Usage Rules	Object Class Term	Properties									
COM000001	HKID Number. Details	1.0	Matured for reuse		ABIE	holder of a Hong Kong Identity Card issued by the Immigration Department of			HKID Number	Details									
COM000002	HKID Number. Identifier	1.0	N/A		BBIE	Hong Kong ID number without check digit			HKID Number	Identifier									
COM000003	HKID Number. Check Digit. Text	1.0	N/A		BBIE	Check digit of Hong Kong ID number			HKID Number	Check Digit									
COM000004	HKSARG Employee. Rank. Code	1.0	Matured for reuse	Rank Code	BBIE	A code representing the rank of a civil servant of the HKSARG.		supplementary component 'Code List Version' to indicate	HKSARG Employee	Rank									
COM000005	HK Business Registration Number. Details	1.0	Matured for reuse		ABIE	A unique reference assigned by the Inland Revenue Department of the HKSARG to a business organization.			HK Business Registration Number	Details									
COM000006	HK Business Registration Number. Business Registration Number. Text	1.0	N/A		BBIE	A unique reference assigned by the Inland Revenue Department of the HKSARG to a business organization			HK Business Registration Number	Business Registration Number									
COM000007	HK Business Registration Number. Branch Number. Text	1.0	N/A		BBIE	Optional Branch Number within the business organization			HK Business Registration Number	Branch Number									
COM000008	Person. Gender. Code	1.0	Matured for reuse		BBIE	A code to indicate the gender of a person (M=male, F=female, U=unknown).		BIDs are recommended to let	Person	Gender									

Figure 4: BIM spreadsheet

BIM provides a modelling spreadsheet as shown in Figure 4 for business analysts to enter the specifications of information models. The spreadsheet contains user-friendly macros that guide users to enter various attributes required by different models on specific dialog boxes. It also allows users to import and select existing models for reuse in defining new models. Moreover, the models specified in the spreadsheet can be converted into XSD code through an XML Stylesheet Language Transformation (XSLT) script without programmers' involvement.

2.1.4 XML schema definition development

The Design Guide provides a comprehensive specification to standardize the conversion each information model into XSD code. For example, each BBIE or ABIE is converted into an XSD complex type. The specification has adopted the Venetian Blind XSD coding convention. A standard library (i.e., XSD file) for the set of CCTs is provided for inclusion in the XSDs developed from the information models. Since the conversion mechanisms, including XSD representations of models as well as the naming convention of types and elements, have been standardized, the conversion from information models to XSD code can be automated. The BIM spreadsheet has provided an XSLT script for this purpose.

2.2 XML schema management guide

The XML Schema Management Guide (Management Guide) defines the framework (e.g., policy, organization structure, etc.) to manage the Project Schemas and Common Schemas defined using the Design Guide.

The Management Guide describes:

- the considerations for the managing Project Schemas as well as their controlled vocabularies, e.g., code lists, namespace, etc.,
- the management process and organization structure for concerted alignment of reusable data elements proposed by B/Ds for standardization of Common Schemas, and
- the guidelines for building a Project Registry and the Central Registry for sharing Project Schemas and Common Schemas respectively.

Each Common Schema is associated with one of the following three maturity levels and the criteria to promote a Common Schema to a higher level are described in Figure 5.

- *Level 0 – Agreed in principle.* B/Ds have generally agreed on the definition, representation, and usage context of the concerned data element, but among those B/Ds that believe the data element might be applicable to their business, the majority anticipate that they require further investigation and analysis before adopting the Common Schema.
- *Level 1 – Recommended for reuse.* B/Ds have generally agreed on the definition, representation, and usage context of the concerned data element, and among those B/Ds that believe the data element might be applicable to their business, the majority anticipate that they are ready to adopt the concerned data element in most of their future projects
- *Level 2 – Matured for reuse.* B/Ds have generally agreed on the definition, representation, and usage context of the concerned data element, and among those B/Ds that believe the data element might be applicable to their business, the majority anticipate that they are ready to adopt the concerned data element in most of their future projects. In addition, the concerned data element has already been used in some projects and the information model of this data element has remained stable for a certain period.

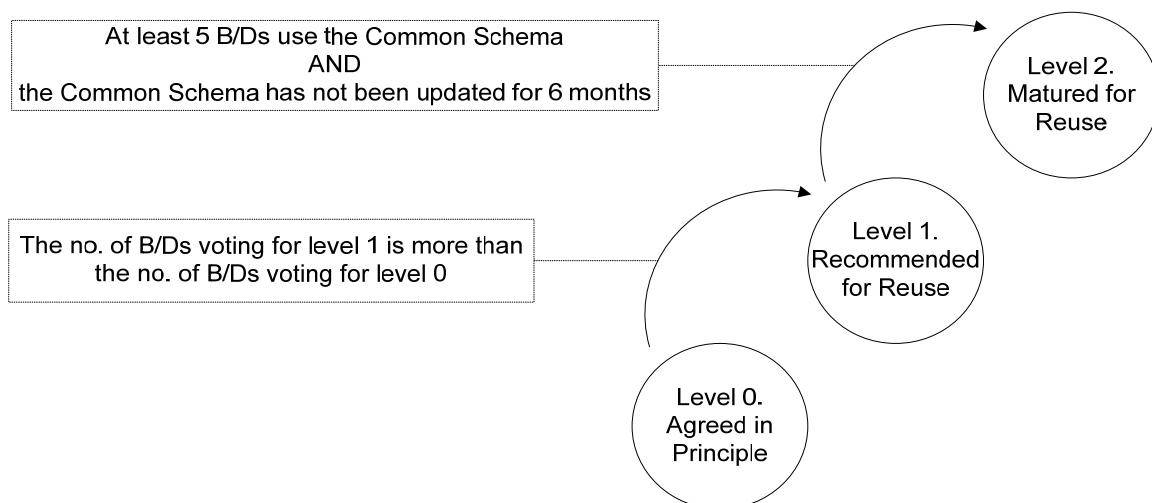


Figure 5: Promotion of the maturity level of a common schema

3. Central registry of data standards

HKSARG has established the Registry of Data Standards (Registry) at <http://www.xml.gov.hk> as the central repository to share the following resources:

- *Common Schemas*, i.e., standardized Common Schemas reusable across B/Ds. Examples of currently available Common Schemas are different address formats, and English and Chinese names of Hong Kong citizens.
- *Project Schemas*, i.e., data models and XML schemas specific to different e-government projects.
- *Code lists and XML namespaces* of Common and Project Schemas.
- *References*, i.e., soft copy of the Guide and all papers discussing the design principles and considerations in aligning data elements.
- *Software tools*, such as modelling spreadsheets (Excel and Open Office), programs to generate XSD code from spreadsheet models, and schema documentation tools.

4. Application in e-government projects

Many e-government projects have applied the Guide to create their Project Schemas. The following are some examples:

- *Automation of dangerous goods manifests submission* (Marine Department 2003). The project facilitates shipping companies to automate submission of dangerous goods (DG) manifests to the HKSARG Marine Department in the form of XML messages through the Internet. The DG messages are transferred using the ebXML Message Service protocol.
- *Weather information in XML* (Hong Kong Observatory 2007). The Hong Kong Observatory publishes weather information in XML for businesses, such as media and transport companies, and government departments to automate processing of weather information.
- *Works Project Information Standard* (Development Bureau 2009). The HKSARG Development Bureau defines a large-scale XML data standard for various building and construction domains to realize electronic data exchange among various stakeholders in construction projects.

5. Clustering service scenario in Macao SAR government

Macao SAR government is pursuing e-government implementation for many years. They have established many e-government facilities, for instance, the Government Portal, public key infrastructure (PKI), Common Payment Gateway, geographic information system (GIS), eForms and many government websites. The current data interoperability standard for Macao SAR government is yet becoming explicitly comprehensive and completed. One important obstacle for such situation is because of the lack of joined-up services among departments, which is the main driving force for reaching data agreements and data alignment standards. Yet, this obstacle does not preclude the establishment of a unified XML schema design and management framework for new e-government scenarios. One important scenario is about how to develop and manage XML schemas for clustering services or shared services. The concept of clustering services was suggested in the six stage e-government transformation model (Turban 2006), which was proposed by Deloitte Research in 2000 and has been widely cited by academic research papers. The Deloitte Six Stage Model provides a comprehensive development path of e-government. This model provides the insight of developing data standards for e-government clustering services. A common service is to bring up and cluster the shared services into a common platform for citizens. Shared services are usually separated from service requesters through the application of information and communication technology. A shared service is a generic service that is jointly developed by government departments and can be used many times in different business processes of various government agencies. Janssen and Wagenaar (2004) suggested that shared services can be developed by unbundling and centralizing activities. Shared services can be developed using Web Services technology. Their architecture should provide the flexibility to include common services and functionality provided by legacy systems, which cannot be replaced easily and would otherwise restrict further development.

In the Deloitte Six Stage Model, the six stages are:

- Stage One: information publishing / dissemination.
- Stage Two: "official" two-way transactions.

- Stage Three: multi-purpose portals.
- Stage Four: portal personalization.
- Stage Five: clustering of common services.
- Stage Six: full integration and enterprise transformation.

In stage five, government will identify shared service and cluster its delivery to citizen so that citizens could view once-disparate services in the portal. The current Macao e-government services are in this stage and start to identify some shared service or joint-up service. For delivery of shared eGovernment services, data interoperability is the essential component in their system architecture.

5.1 Adopting HKSARG XML schema design and management in Macao

In view of adopting the HKSARG XML Schema Design and Management in Macao, a unified framework to design standardized Common Schemas for e-government clustering services or shared services receives high attention due to its practicability. In the adopted method for Common Schema development, a clustering service will be nominated and selected by an e-government task force. The nomination of the shared service is based on the consideration of its commonness and separability from the services being delivered. For instance, the address change service is a typical shared service among different government departments. The address data is common in most government services and its administrative functions could be reorganized as a shared service, e.g., for creation, update, deletion and authorization of address data. To turn the address change service (ACS) into a shared service, the main considerations heavily rely on the data interoperability on the address data exchanged among government departments. This requires an address data standard for ACS. During the system design of the nominated clustering service, business process and information modelling will be done. In the ACS case, the workflows and the computerized forms for address creation and modification will be captured, analyzed, and modelled using BPM and BIM spreadsheets. Then, the macro program in BIM spreadsheet is run to generate the required XSD code from the data model. For the XML schema management framework, since clustering services are the common among the majority of the government departments, the developed Common Schema would enter directly into maturity level 1 and become the recommended reusable data schema for future projects.

6. Conclusions

The XML Schema Design and Management Guide provides a comprehensive XML schema design methodology and the necessary schema management infrastructure to facilitate e-government data standardization. A library of reusable Common Schemas has been developed, and shared in a data standards registry for developing Project Schemas for new government services. While these schemas are designed for government projects, they are also published publicly to facilitate adoption by different business sectors for development industry-specific data standards. Moreover, the XML Schema Design Guide is a generic integrated methodology, which covers all necessary steps in e-business modelling, from requirement analysis, to business process and information modelling, to XSD coding. Therefore, business enterprises and other governments can easily adapt it to establish their own data interoperability infrastructures.

References

- Development Bureau (2009), HKSARG, *Works Project Information Standard*, [Online], Available: http://www.devb-wb.gov.hk/electronic_services/works_project_info/index.aspx?langno=1&nodeID=1675.
- HKSARG (2008), *Digital 21 Strategy*, [Online], Available: <http://www.info.gov.hk/digital21>.
- HKSARG (2009), *The Registry of Data Standards*, [Online], Available: <http://www.xml.gov.hk>.
- Hong Kong Observatory (2007), HKSARG, *Weather Information in Extensible Markup Language (XML)*, [Online], Available: <http://demo.xml.weather.gov.hk>.
- Janssen, M., R. Wagenaar (2004), *Developing Generic Shared Services for e-Government*, Electronic Journal of e-Government Volume, 2004. 2(1): p. 31-38.
- Layne, K., J. Lee (2001), *Developing fully functional E-government: A four stage model*, Government Information Quarterly.
- Marine Department (2003), HKSARG, *XMLDG Project Website*, [Online], Available: <http://ebs.mardep.gov.hk/xmldg/>.
- OGCIO (2008), HKSARG. *Interoperability Framework for E-Government*, [Online], Available : <http://www.ogcio.gov.hk/eng/infra/eif.htm>.
- OGCIO (2006), HKSARG, *XML Schema Design and Management Guide v1.3*, [Online], Available : <http://www.ogcio.gov.hk/eng/infra/eif.htm>.

OASIS (2001), *ebXML Business Process 1.01*, [Online], Available:

OASIS (2008), *Universal Business Language*, [Online], Available: <http://www.oasis-open.org/committees/ubl>.

Silcock, R. (2001), *What is E-government*. Parliamentary Affairs 2001, 54(1): p. 88-101.

Turban, E (2006). *Electronic Commerce: A Managerial Perspective*, Prentice Hall.

UN/CEFACT (2001), *UMM Revision 10*, [Online], Available: http://www.unece.org/cefact/umm/umm_index.htm.

UN/CEFACT (2003), *ebXML Core Components Technical Specification 2.01*. [Online].

Available: http://www.unece.org/cefact/ebxml/CCTS_V2-01_Final.pdf.

W3C (2004), *XML Schema 1.1*, [Online], Available: <http://www.w3.org/XML/Schema>.

W3C (2006), *Extensible Markup Language (XML) 1.1*, [Online], Available: <http://www.w3.org/XML>.