2008

A Data Flow Perspective for Business Process Integration

Xitong (Maxwell) Guo
University of Science and Technology of China/City University of Hong Kong, maxwell@mail.ustc.edu.cn

Sherry X. Sun
City University of Hong Kong, sherry.sun@cityu.edu.hk

Doug Vogel
City University of Hong Kong, isdoug@cityu.edu.hk

A DATA FLOW PERSPECTIVE FOR BUSINESS PROCESS INTEGRATION

L'intégration des processus d'affaires selon une perspective de flux de données

Research-in-Progress

Xitong (Maxwell) Guo
University of Science and Technology of China-City University of Hong Kong
Joint Research Center,
Suzhou, P.R. China
maxwell@mail.ustc.edu.cn

Sherry X. Sun
Department of Information Systems,
City University of Hong Kong,
83 Tat Chee Ave
sherry.sun@cityu.edu.hk

Doug Vogel
Department of Information Systems,
City University of Hong Kong,
83 Tat Chee Ave
isdoug@cityu.edu.hk

Abstract

Business process integration has become prevalent as business is increasingly crossing organizational boundaries. While workflow technology is a standard solution for business process management, it is imperative for workflow management systems to provide effective and efficient support for collaboration. To address the issue of protecting organizations' competitive knowledge and private information while also enabling business-to-business (B2B) collaboration, past research has focused on customized public and private process design and structure correctness of the integrated workflow. However, data flow is important for business process integration because data is always sensitive when conducting inter-organizational business and data errors could still happen even given syntactically correct activity dependence. This paper presents a data flow perspective. It gives an approach to define a "public data set" for each involved organization exemplifying the integrated workflow that is needed in order to be free from data anomalies e.g., missing data and redundant data errors.

Keywords: Data flow, inter-organizational data errors, workflow, business process integration

Résumé

Dans une objectif d'intégration des processus d'affaires, la recherche passée s'est concentrée sur la personnalisation du processus aux besoins du client et sur l'exactitude de conception et de structure des processus interne. Dans le contexte de processus inter-organisationnels, cet article retient une perspective en termes de workflow. Il propose de définir une « base de données publiques » pour chaque organisation impliquée en exemplifiant le workflow requis afin d'éviter la présence d'anomalies dans les données.

Introduction

Organizations increasingly find themselves engaging in interactions with other organizations as well as local and global government entities in circumstances where there is no single point of authority and control. These multiple independent stakeholders circumstances require special attention be paid to the management of processes. Workflow technology has become a standard solution to enhance effectiveness and efficiency for managing complex processes in businesses domains such as supply chain management, E-commerce, customer relationship management, and knowledge management (Bussler 2002; Stohr and Zhao 2001; Sun et al. 2006).

In current literature, several issues are considered when doing business process integration: structural correctness, privacy and knowledge protection, flexibility, and workflow standards for inter-organizational collaboration. Structural correctness and privacy protection are the two fundamental issues to ensure that the integrated process is robust and secure(Schulz & Orlowska 2004). In the aspect of structure correctness, the Petri-net-based approach is used to analyze and verify syntactic errors in control flow of the integrated workflows. It is also used to describe the concepts of workflow merging, and basic properties that an inter-organizational workflow should satisfy in order to achieve soundness (Shuang et al. 2006; van der Aalst 2000). In the aspect of privacy protection, design of customized process for each partner is most studied. On the one hand, cooperation needs a certain degree of workflow inter-visibility in order to perform collaboration. On the other hand, organizations need privacy to protect their competency (organizations desire only selected partners of process to be visible). So the question is how to balance between "public" and "private" processes. Public-To-Private (P2P) approach, customized process views and B2B protocols are described to address this issue for inter-organizational workflows (Bussler 2002; Eshuis and Grefen 2008; Liu et al. 2006; van der Aalst and Weske 2001). The literature focuses on other issues, such as the collaboration standards is not listed here, e.g., eXchangeable Routing Language (van der Aalst and Kumar 2003). Previous studies mainly focus on customized public and private processes and control flow dependencies between activities. However, a data flow perspective is important for business process integration as most activities in a process need to read, write, and alter data. As such, workflow engines have to be supported by databases.

Two challenges have motivated us to address the data flow perspective for business process integration. First, data is always sensitive when conducting inter-organizational business. B2B relationships require the sharing of information and they are managed using cooperative and competitive postures (Klein et al. 2007). A key question for each party is how much information (i.e., retailer's point-of-sale data, supplier's capacity data) to share with partners to achieve outcome equilibrium (Lee et al. 2000; Cachon & Lariviere 1999). In addition to the benefits of a broader range of information sharing, there is also reluctance to share due to the confidential nature of some information (Gosain et al.2004; Sawaya 2006). For the contradicting requirement of public visibility versus privacy, an important question is, what is the data set that should be exposed precisely as the integrated process requires, yet still keeping data privacy? Second, while the integrated processes have a syntactically correct process sequence, they cannot execute collaboratively if the interface dialogue does not support the correct data flow. It is worth noting that data flow errors can happen not only in intra-organizational workflows (Sun et al. 2006), but also in inter-organizational workflows where data is transferred across organizational boundaries. This kind of error cannot be detected by the control flow analysis approach. And these data errors (such as missing data errors and redundant data errors) always involve high costs to figure out and to fix across organizations. Then, what is the condition that the integrated workflow should satisfy in order to be free from these data errors?

For the above two questions, current literature and commercial systems only give scant treatment. Past studies about B2B integration (e.g., P2P approach, customized process views), are more focused on structural correctness based on a control flow approach. Commercial systems are mostly visualization tools with few analytical capabilities. Standards such as RosettaNet define domain-specific public processes called Partner Interface Processes (PIPs), but do not offer analytical capability. This paper provides a data flow perspective for business process integration to define the "public data set" precisely for each involved partner as the integrated workflow requires in order to be free from data errors. We present a theoretical framework for representation and analysis of data flow for business process integration.

A Business Process Integration Example

In this section, we give a business example in the chained execution forms. Three organizations, Factory, Third-party Logistics and Customs are involved in the business scenarios. "Factory" focuses on the core business of production and outsources the "pick & pack" while "customs declaration" services to Third-party Logistics. Customs executes the process of "customs declaration approval." Figure 1 illustrates the integrated Factory_Logistics_Customs global process, the business logic is Factory firstly executes "production", then sends an order to Logistics, Logistics receives the order, evaluates it and decides whether to reject it or not. If the order is rejected, Factory is notified; otherwise, pick & pack is executed and customs declaration application is sent to customs. After the customs declaration approval process, Customs send notification to Logistics, and Logistics deliver it to factory.

In Figure 1, Partner Interface Processes (PIPs) are created to provide an interface between two different enterprise processes (El Sawy 2001). The PIP is comprised of the "touch-point" activities from the two different enterprise processes and the dialogue among those activities via data exchange. The boxes in bold are PIPs that transfer data between partners, boxes in blank are each party's private activities, which are kept as secrets or not interested by partners. For example, Factory's PIPs include "Send order", "Receive rejection", "Receive notification", these are public activities to exchange data with Third-party Logistics.

Logistics and Customs individual processes and data flow are shown in Figures 2a, Figure2b and Figure 3a, Figure 3b respectively. Figure 2b and Figure 3b use UML extended with data flow to model workflow (for detail please refer to Sun et al. (2006)). Each activity in the data diagram has an input and output data set, denoted as *I* and *O*, respectively. As the factory individual process is simple, we do not show it here. Its private activity "Production" v_0, $I_{v_0} = \phi$, $O_{v_0} = \{d_1, d_2, d_3, d_4, d_5\}$. Symbols used in the local process and global process are listed in Table 1.

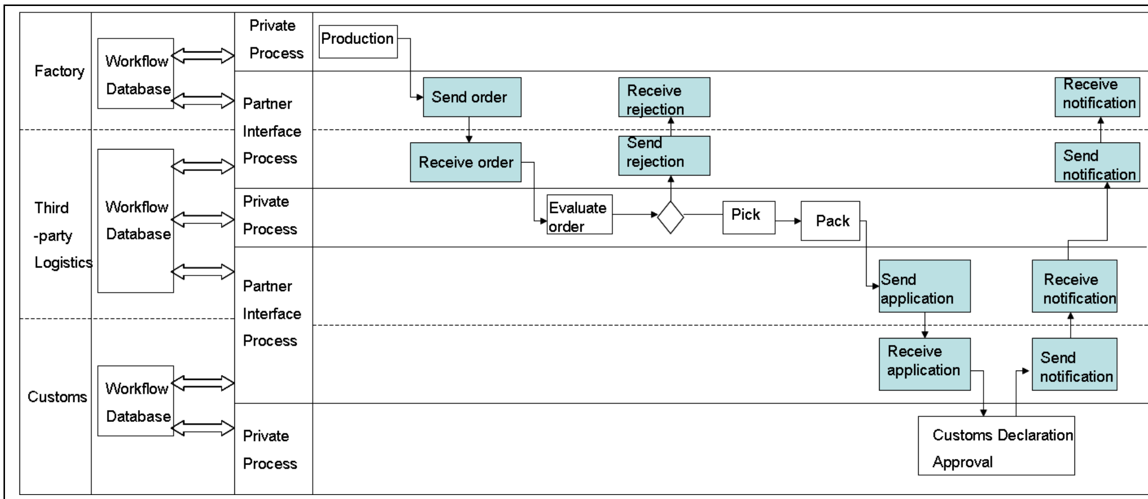

Figure 1. Factory_Logistics_Customs Integrated Workflow

Table 1. Symbols Used in Local Processes & Factory_Logistics_Customs Integrated Process

Data items		Activities	
d_1 Goods name & coding	d_{12} Container estimate value	v_0 Factory production	v_{12} Decision node 2
d_2 Goods weight	d_{13} Trade permit	v_1 Factory send order to logistics	v_{13} Request missing information
d_3 Goods volume	d_{14} Exporter copy form	v_2 Logistics receive order	v_{14} Check goods quality
d_4 Goods quality certificate	d_{15} Invoice	v_3 Logistics evaluate order	v_{15} Check goods dangerous level
d_5 Goods estimate value	d_{16} Application summary	v_4 Decision node 1	v_{16} Decision node 3

Data items		Activities	
d_{6a} Goods destination port	d_{17} Application complete	v_5 Logistics notify factory that they reject the order	v_{17} Forward to assessing officer for signature
d_{6b} Container destination port	d_{18} Quality verified	v_6 Factory receive order rejection	v_{18} Decision node 4
d_7 Pick summary	d_{19} Dangerous level verified	v_7 Logistics pick goods from factory	v_{19} Forward commissioner for signature
d_8 Order doable	d_{20} Signed by assessing officer	v_8 Logistics pack goods	v_{20} Customs send notification to Logistics
d_9 Packing list	d_{21} Signed by commissioner	v_9 Logistics sends customs declaration application to Customs	v_{21} Logistics receive notification
d_{10} Container weight	d_{22} Customs declaration rejection	v_{10} Customs receive application	v_{22} Logistics send notification to factory
d_{11} Container amount		v_{11} Verify completeness of application	v_{23} Factory receive notification

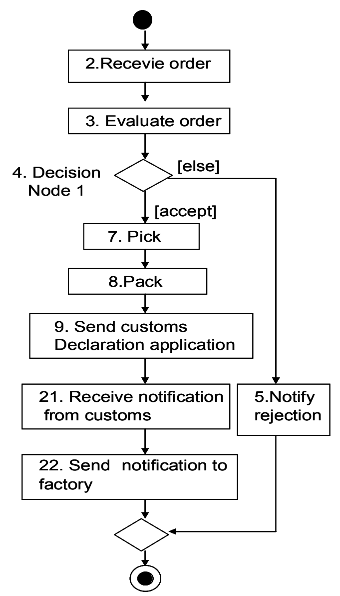

Figure 2a. Logistics Process

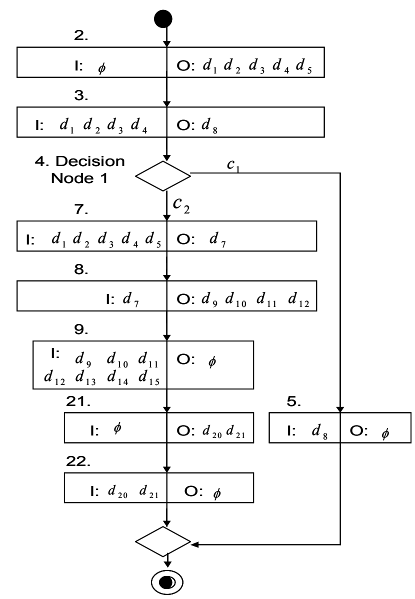

Figure 2b. Data Flow in the Logistics Process

Table 2. Symbols Used in Concepts of Workflow Integration	
V a set of activities	I_i^e the set of external data to v_i
W, W_a, W_b, W_g Workflow model	I^e the set of external data to W, $I^e = \bigcup I_i^e = I^{el} \bigcup I^{eO}$
C_W the routing constraint set for W	I^{eO} the set of external data to W, and not available from local organization
Γ_W all the instance set of W	I^{el} the set of external data to W, but is available from local organization
Γ, Γ_i any instance set	I^{eOA} The outside_external_data set that organization A needs for private activities set
I_i data set as input for activity v_i	Γ_{b_Pri} Private activities set belong to organization b

Concepts of Workflow Integration with Data Flow

This section gives concepts in data flow and presents two new definitions that are later used in data flow analysis for workflow integration. Symbols used are given in table 2. We begin by review some concepts in a single workflow.

Workflow Instance set & Constraint set A set of activities is executed in a specified order from the start activity to the end, requiring a set of constraints $C = C_W$ to be satisfied. The set of the activities is workflow instance, denoted

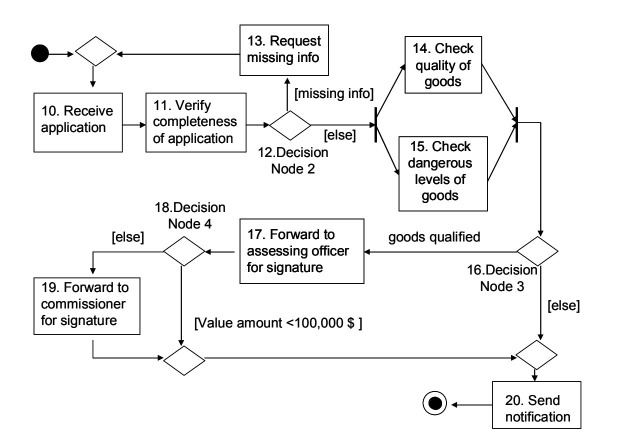

Figure 3a. Customs Declaration Approval Process

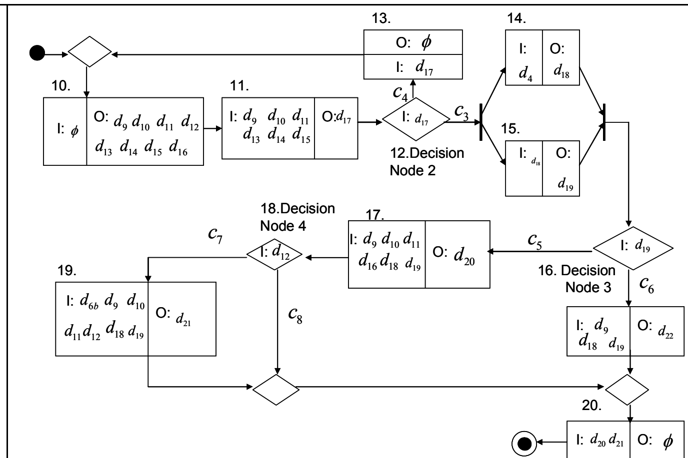

Figure 3b. Data Flow in the Customs Declaration Approval Process

as Γ . The corresponding set C is called the routing constraint set of Γ . Detailed definition is available from Sun et al. (2006). For example, a workflow instance for the Third-party Logistics local workflow is $\Gamma_1 = \{v_2, v_3, v_4, v_5, v_6\}$, corresponding routing constraint set C= {the data item "order doable" d_8 = No}.

Next we give two definitions relate to workflow integration with data flow.

Definition 1 (Workflow model with data flow) When a workflow is denoted as $W(V,\Delta)$, V is a finite activity set including all the possible activities executed in workflow $V = \{v_1, v_2, ...v_m\}$, m is the total account of activities, Δ is the data set that all activities in the workflow read and write, $\Delta = \bigcup_{i=1}^{m}(I_i \bigcup O_i)$.

For example, the Logistics individual workflow $W_L(V_L, \Delta_L)$, $V_L = \{v_2, v_3, v_5, v_7, v_8, v_9, v_{21}, v_{22}\}$,
$\Delta_L = \left\{\bigcup_{i=1}^{5} d_i, d_{6a}, d_{6b}, \bigcup_{i=7}^{15} d_i, d_{20}, d_{21}, d_{22}\right\}$.

Definition 2 (Workflow integration) Workflow W_a (V_a, Δ_a) and workflow W_b (V_b, Δ_b) owned by organizations A and B, they are integrated and create the global workflow $W_g(V_g, \Delta_g)$ where $V_g = V_a \bigcup V_b$, $\Delta_g = \Delta_a \bigcup \Delta_b$. $\Gamma_{W_a}, \Gamma_{W_b}$ are all the instance sets in W_a and W_b, $\Gamma_g = \Gamma_{W_a} \times \Gamma_{W_b}$; C_{W_a}, C_{W_b} are all the instance sets in W_a and W_b, $C_{W_g} = C_{Wa} \times C_{W_b}$.[1]

In the Factory_Logistics_Customs global workflow, there is $\Gamma_1 \in \Gamma_g$,

$$\Gamma_1 = \{v_0, v_1, v_2, v_3, v_4, v_7, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}, v_{21}, v_{22}, v_{23}\}$$

$$= \{v_0, v_1, v_{23}\} \bigcup \{v_2, v_3, v_4, v_7, v_8, v_9, v_{21}, v_{22}\} \bigcup \{v_{10}, v_{11}, v_{12}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{20}\}$$, with constraint set

C={ d_8 = Yes, d_{17} = Yes, d_{18} = Yes, d_{19} = Yes, $d_{12} \geq$ 100,000}

---

1 $V_g = V_a \bigcup V_b$, it is lossless integration; else, $\exists v \in V_a \bigcup V_b, v \notin V_g$, i.e. $V_g \subset V_a \bigcup V_b$ it is lossy integration. This is the same with the definition in Shuang et al. (2006) with Petri net approach. Here we do not take lossy integration into consideration.

Data Flow Analysis for Workflow Integration

This section shows a data flow analysis approach for workflow integration based on the data flow framework in a single workflow. First we review data errors in a single workflow. For more detail please refer to Sun et al. (2006).

Missing data error When an activity v in workflow needs data d as input data, d has not yet been initialized, the missing data error happens. As data is initialized either as an output of some activity or as external input item, we can see that the workflow is free from missing data error only if

$\forall \Gamma_i \in \Gamma_w$, $\bigcup_{i=1}^{m} I_i \subseteq (\bigcup_{i=1}^{m} O_i) \bigcup I^e$ holds. Condition (1)

In the Factory-Logistics-Customs integrated global workflow $W_g\left(V_g, \Delta_g\right)$, there is $\Gamma = (\bigcup_{i=0}^{23} v_i) \setminus \{v_5, v_6\}$, with constraint set $C_{g2} = \{c_2, c_3, c_5, c_7\}$, external data item d_6 "Container destination port" is needed as an input for activity v_{19} "Forward to commissioner for signature", but it is not transferred to Customs. This is an example of missing data error.

Redundant data error In a workflow $W(V, \Delta)$, when a data item d is input as an external data or produced by an activity $v \in V$, but d is not consumed by any activity in W as input data item, the redundant data error happens.

In the Factory-Logistics-Customs integrated global workflow $W_g\left(V_g, \Delta_g\right)$, there is $\Gamma \in \Gamma_{W_g}$, $\Gamma = \{v_1, v_2, v_3, v_5, v_{21}\}$ with corresponding routing constraint sets $\{c_1\}$, external data item d_5 "Goods estimate value" is not used by any activity in Γ. This is an example of redundant data error.

According to the definition, the workflow is free from redundant data error only if

$\forall \Gamma_i \in \Gamma_w$, $((\bigcup_{i=1}^{m} O_i) \bigcup I^e) \setminus (\bigcup_{i=1}^{m} I_i) = \phi$ holds. Condition (2)

Based on the above two conditions, we get

Lemma 1 The workflow is free from missing data error and redundant data error only if

$\forall \Gamma_i \in \Gamma_w$, $((\bigcup_{i=1}^{m} O_i) \bigcup I^e) = \bigcup_{i=1}^{m} I_i$ holds. Condition (3)

Lemma 2 If Condition (4) "$\forall \Gamma_g \in \Gamma_{W_g}$, $((\bigcup_{i=1}^{m} O_i) \bigcup I^e) \setminus (\bigcup_{i=1}^{m} I_i) = \phi$" holds, then

$\forall \Gamma_a \in \Gamma_{W_a}$, $((\bigcup_{i=1}^{m} O_i) \bigcup I^e) = \bigcup_{i=1}^{m} I_i$ holds and Condition (5)

$\forall \Gamma_b \in \Gamma_{W_b}$ $((\bigcup_{i=1}^{m} O_i) \bigcup I^e) = \bigcup_{i=1}^{m} I_i$ holds. Condition (6)

Condition (4) is at the global workflow level. However, as $\Gamma_g = \Gamma_{W_a} \times \Gamma_{W_b}$, we turn the goal to the individual workflow level.

Given $W_a\left(V_a, \Delta_a\right)$, $W_b(V_b, \Delta_b)$ and global workflow $W_g\left(V_g, \Delta_g\right)$

Let Δ_{A_Pub} be the data set organization A publicize and transferred to organization B in instance $\Gamma \in \Gamma_g$,

I^{eOA} be the external data set that organization A needs and have to get from organization B in $\Gamma \in \Gamma_g$,

I^{eIA} be the external data set that organization A needs but are available from local organization in $\Gamma \in \Gamma_g$.

Δ_{B_Pub}, I^{eOB}, I^{eIB} are the symmetric definition for organization B.

Proposition 1 $\forall \Gamma_g \in \Gamma_{W_g}, \Gamma_g = \{\Gamma_a, \Gamma_b\}, \Gamma_{a\_Pri} \in \Gamma_a \in \Gamma_{W_a}, \Gamma_{b\_Pri} \in \Gamma_b \in \Gamma_{W_b}$, Γ_{a_Pri} and Γ_{b_Pri} are private processes belong to organization A and B respectively, in order for the two goals

1) To eliminate missing data errors and redundant data errors in the global workflow.

2) Each involved organization exposes as less data as required by the collaboration.

$$\Delta_{A\_Pub} = I^{eOB},\ \Delta_{B\_Pub} = I^{eOA} \quad \text{holds}$$

Where $I^{eOB} = (\bigcup_{i=1}^{m} I_i) \setminus ((\bigcup_{i=1}^{m} O_i) \bigcup I^{eIB})$, m is the total account of activity in Γ_{b_Pri}

$I^{eOA} = (\bigcup_{i=1}^{m} I_i) \setminus ((\bigcup_{i=1}^{n} O_i) \bigcup I^{eIA})$, n is the total account of activity in Γ_{a_Pri} .

Proof: Figure 4 shows the data flow in public and private processes. Global workflow instance $\Gamma_g = \{\Gamma_a, \Gamma_b\} = \{\Gamma_{a\_Pri}, \Gamma_{a\_Pub}, \Gamma_{b\_Pri}, \Gamma_{b\_Pub}\}$. Through public activities in Γ_{a_Pub} and Γ_{b_Pub} , organization B exposes data set Δ_{B_Pub} to organization A, and Δ_{B_Pub} are passed to activities in Γ_{a_Pri} .

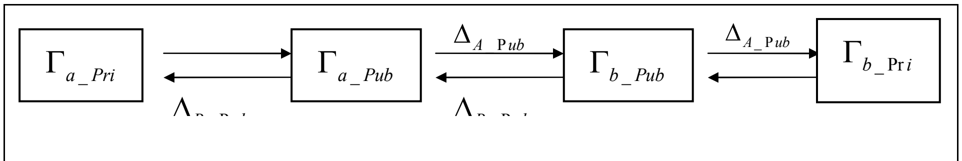

Figure 4. Data Flow in Public & Private Processes

For goal 1), according to Lemma 2, we get $((\bigcup_{i=1}^{m} O_i) \bigcup I^e) = \bigcup_{i=1}^{m} I_i$ for Γ_{a_Pri} . Here $I^e = I^{eOA} \bigcup I^{eIA}$.

As I^{eOA} is the external data set that are not available within Organization A, if $\exists d \in O_i$, there is no need to ask data item d from the partner, that is $I^{eOA} \bigcap \bigcup_{i=1}^{m} O_i = \phi$. Obviously, $I^{eOA} \bigcap I^{eIA} = \phi$.

Then, $I^{eOA} = (\bigcup_{i=1}^{m} I_i) \setminus ((\bigcup_{i=1}^{n} O_i) \bigcup I^{eIA})$.

For goal 2) , $\Delta_{B\_Pub} \subseteq I^{eOA}$. If $\exists d \in I^{eOA}, d \notin \Delta_{B\_Pub}$, a missing data error happens.

Thus, $\Delta_{B\_Pub} = I^{eOA} = (\bigcup_{i=1}^{m} I_i) \setminus ((\bigcup_{i=1}^{n} O_i) \bigcup I^{eIA})$. And the same logic for Δ_{A_Pub} .

Public data set determination based on reverse business logic and activity dependence order[2]

Here we present an approach to determine the public data set for each involved organization based on Proposition 1.

Business is executed in order, for example, the Factory_Logistics_Customs integrated process execution order is Factory-> Logistics-> Customs.

Activity has the relationship of dependence, denoted as $\prec$. Activity $v_i \prec v_j$ when $O_i \bigcap I_j \neq \phi$ or $\exists v_k, v_i \prec v_k \prec v_j$. In Logsitics local workflow, $O_{v_7} \bigcap I_{v_8} = \{d_7\}$, thus, $v_7 \prec v_8$. The business rule is " $d_{6a} \in I_{v_7}$ only if $d_{6b} \in O_{v_8}$ ". That is, only when the activity "Pack" has output data item "Container destination port" d_{6b} , activity "Pick" needs data "Goods destination port" d_{6a} as input.

We use reverse business logic and activity dependence order to determine public data set.

Take a workflow instance Γ_g in the Factory_Logistics_Customs integrated global workflow as an example.

---

[2] The general principle of public data set can be detailed through an algorithm, which is available upon request.

Step 1 An instance in a global workflow is combined by local workflow instances, and each instance has private and public activity sets.

$\Gamma_g \in \Gamma_{W_g}$, $C_{W_g} = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8,\}$, $c_1 = \{d_8 = No\}$, $c_2 = \{d_8 = Yes\}$, $c_3 = \{d_{17} = Yes\}$, $c_4 = \{d_{17} = No\}$, $c_5 = \{d_{18} = Yes, d_{19} = Yes\}$ $c_6 = \{d_{18} = No, or, d_{19} = No\}$, $c_7 = \{d_{12} \geq 100{,}000\}$, $c_8 = \{d_{12} < 100{,}000\}$.

$\Gamma_g = (\bigcup_{i=0}^{23} v_i) \setminus \{v_5, v_6, v_{13}\}$, with constraint set $C_g = \{c_2, c_3, c_5, c_7\}$, c_7 is triggered when container estimate value is high.

Factory's private activity set $\Gamma_{F\_Pri} = \{v_0\}$, Logistics' private activity set $\Gamma_{L\_Pri} = \{v_3, v_4, v_7, v_8\}$

Custom' private activity set $\Gamma_{C\_Pri} = \{v_{11}, v_{12}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}\}$.

Step 2 Calculate outsite_external_data for each involved organization.

Rule I based on reverse business logic.

Firstly, pick Customs. Custom's outsite_external_data set

$$I^{eOC} = \{\bigcup_{i=11}^{19} I_i\} \setminus \{\bigcup_{i=11}^{19} O_i, \bigcup_{i=11}^{19} I_i^{eIL}\} = \{d_4, d_{6b}, d_9, d_{10,} d_{11}, d_{12}, d_{13}, d_{14}, d_{15}\}$$

Thus, Logistics' public data set transferred to Customs $\Delta_{L\_Pub} = I^{eOC} = \{d_4, d_{6b}, d_9, d_{10,} d_{11}, d_{12}, d_{13}, d_{14}, d_{15}\}$.

Rule II based on reverse activity dependence order.

Secondly, pick Logistics. After determine Δ_{L_Pub} , we know that data item d_{6b} "Container destination port" is needed to be passed to Customs from Logistics. For $\Gamma_{L\_Pri} = \{v_3, v_4, v_7, v_8\}$, $v_7 \prec v_8$. As $d_{6b} \in \Delta_{L\_Pub}$, then $d_{6b} \in O_{v_8}$, and we decide that activity v_7 "Pick" needs d_{6a} as input. Then, Logistics' outsite_external_data set

$$I^{eOL} = \{I_3, I_4, I_7, I_8\} \setminus \{O_3, O_4, O_7, O_8, I_3^{eIL}, I_4^{eIL}, I_7^{eIL} I_8^{eIL}\} = \{d_1, d_2, d_3, d_4, d_5, d_{6a}\}$$

And Factory's public data set transferred to Logistics $\Delta_{F\_Pub} = I^{eOL} = \{d_1, d_2, d_3, d_4, d_5, d_{6a}\}$.

Thus, we get the Factory's and Logistics' public data set Δ_{F_Pub} , Δ_{L_Pub} .

Discussions and Future Research

This paper studies business process integration from a data flow perspective. Based on fundamental concepts, we give a formal analysis approach to define a "public data set" for each involved organization exemplifying the integrated workflow that is needed in order to be free from data anomalies e.g., missing data and redundant data errors. Our work is promising for decision makers to perform simulations to decide the best data set to expose to a partner. Virtual enterprises are formed when companies collaborate with each other in a virtual way to deal with highly agile market. They are instant and dynamic. For different instances, they transfer different data set to partners.

This study is but a first step and research opportunities abound.

1. It is important to develop a data flow negotiation protocol between business partners. Global workflow instances are collectively controlled by each partner's routing constraints. More importantly, every stakeholder's public data set is determined by the partner's request. How to sense the partner's needs and to communicate its own needs to the partners is an important research question (Gosain et al.2004). Also, data vary in the degree of privacy and security. How should each party deal with this variation in their competitive and cooperative environment? A negation and signal mechanism between business partners will be of much help for each stakeholder to publicize exactly the right data for collaboration.

2. Another challenge is identifying and analyzing significant processes responsible for possible data errors. Given a complex integrated workflow, some processes could be responsible for data errors in specific instances. For instance, in our example the activity "send customs declaration application" always leads to missing/redundant data errors in the integrated processes. Thus, a formal solution to identify and manipulate these processes is important.

References

Bussler, C. "The role of B2B protocols in Inter-Enterprise Process Execution," *in Proceedings of the Workshop on Technologies for E-Services,* Rome, Italy, 2001, pp. 16-29.

Bussler, C. "The Application of Workflow Technology in Semantic B2B Integration," *Distributed and Parallel Databases* (12:2), 2002, pp. 163-191.

Cachon, G.P., and Lariviere, M.A. "Capacity Choice and Allocation: Strategic Behavior and Supply Chain Performance," *Management Science* (45:8), 1999, pp. 1091-1108.

Chebbi, I., Dustdar S., and Tata S. "The view-based approach to dynamic inter-organizational workflow cooperation," *Data & Knowledge Engineering* (56:2), 2006, pp. 139-173.

El Sawy, O. A. *Redesigning Enterprise Processes for E-Business,* McGraw-Hill, Inc. New York, NY, USA, 2001.

Eshuis, R., and Grefen P. "Constructing customized process views," *Data & Knowledge Engineering* (64:2), 2008, pp. 419-438.

Gosain, S., Malhotra, A., and El Sawy, O.A. "Coordinating for Flexibility in e-Business Supply Chains," *Journal of Management Information Systems* (21:3), 2004, pp. 7-45.

Klein, R., Rai, A., and Straub, D.W. "Competitive and Cooperative Positioning in Supply Chain Logistics Relationships," *Decision Sciences* (38:4), 2007, pp. 611-646.

Lee, H.L., So, K.C., and Tang, C.S. "The Value of Information Sharing in a Two-level Supply Chain," *Management Science* (45:5), 2000, pp. 626-643.

Liu, D.R., and Shen, M. "Facilitating Cross-organisational Workflows with a Workflow View Approach," *Decision Support Systems* (38:3), 2004, pp. 399-419.

Sawaya III, W. J. "An investigation of the performance impact of the extent of inter-organizational information sharing: Using a complex adaptive system paradigm and agent-based simulation," Ph.D. Dissertation, University of Minnesota, 2006.

Sayal, M., Casati, F., Dayal, U., and Shan, M.C. "Integrating Workflow Management Systems with Business-to-Business Interaction Standards," in *Proceedings of the 18th International Conference on Data Engineering,* 2002, pp. 287-296.

Schulz, K. A. and Orlowska, M. E. "Facilitating cross-organisational workflows with a workflow view approach," *Data & Knowledge Engineering* (51:1), 2004, pp. 109-147.

Sun, S., Kumar A., and Yen, J. "Merging workflows: A new perspective on connecting business processes," *Decision Support Systems* (42:2), 2006, pp. 844-858.

Stohr, E. A. and Zhao, J. L. "Workflow Automation: Overview and Research Issues," *Information Systems Frontiers* (3:3), 2001, pp. 281-296.

Sun, S.X., Zhao, J.L., Nunamaker, J.F., and Sheng, O.R.L. "Formulating the Data-Flow Perspective for Business Process Management," *Information Systems Research* (17:4), 2006, pp. 374-391.

van der Aalst, W. "Loosely coupled interorganizational workflows: modeling and analyzing workflows crossing organizational boundaries," *Information & Management* (37:2), 2000, pp. 67-75.

van der Aalst, W. M. P. and Kumar A. "XML-Based Schema Definition for Support of Interorganizational Workflow," *Information Systems Research* (14:1), 2003, pp. 23-46.

van der Aalst, W. M. P. and Weske M. "The P2P Approach to Interorganizational Workflows, " In *Proceedings of the 13th International Conference on Advanced Information Systems Engineering,* Interlaken, Switzerland, 2001, pp. 140-156.