Important Issues for Evaluating Inter-Organizational Data Integration Configurations

Frank G. Goethals
IESEG School of Management, Lille, France
F.Goethals@ieseg.fr

Abstract: Partnering companies can share data via various configurations. Typically problems become evident as partnering companies seek to exchange data. These problems take a different form for different data integration configurations and are thus of great relevance when evaluating different configurations. This paper identifies issues to be taken into account when evaluating inter-organizational data integration configurations. Eight problems are discussed: difficulties to identify which information flows to automate; to relate different viewpoints on boundary objects; to agree on data formats; to distribute investments among the parties; to deliver appropriate service levels; to preserve the value of the data sharing; to clarify data ownership and to do all of this in the frame of changing relationships. For several problems it is illustrated how they surface in a completely centralized and in a completely decentralized inter-organizational data-integration scenario.

Keywords: business-to-business integration, inter-organizational data integration problems, boundary objects, service levels, data ownership, data functionality

1. Introduction

Optimization efforts often focus on inter-organizational activities. Such schemes heavily rely on information sharing for two principal reasons: (1) as information is being shared, data inconsistencies across enterprises can be eliminated so that all companies accurately perceive the current state and (2) as new information is being shared, new business practices become possible.

The process-paradigm indicates companies can execute a task if they are in a specific state. Information and Communication Technology (ICT) makes it possible to transmit information about a multitude of states in real-time. For example, in the past only two states were recognized in the ordering process: ‘order placed’ and ‘delivery received’. Nowadays a customer company can be informed that an order was received, accepted, scheduled, picked, loaded in the truck, etc.

A number of ‘information sharing problems’ are associated with setting up an inter-organizational system to share data. Section 2 identifies eight problems and illustrates them in the context of the following practices:

- Vendor Managed Inventory (VMI). The Bullwhip Effect is a significant problem that is discussed extensively in Supply Chain Management literature. Variability in demand is magnified at each stage up the supply chain (i.e., from reseller, over seller to manufacturer). Case studies have proven that, through Vendor Managed Inventory, the Bullwhip-effect can be strongly reduced (Smaros et al. 2003). VMI requires an intensive sharing of stock and sales data among different companies in the supply chain. Information exchange in this configuration is so extensive that it is unrealistic to assume the data could be shared and processed manually, without directly connecting the computer systems of the different companies.
- Product Lifecycle Management (PLM). According to CIMdata (2005), PLM involves the collaborative creation, management, dissemination and use of product definition information across the Extended Enterprise from concept to end of life of the product. Takalikar (2004) asserts companies with PLM get better products to market faster and provide better support to the customers.
- - A specific case: e-procurement via Tradcom. Tradcom (Muylle & Croon 2003) offers a virtual trading environment for companies in Belgium, the Netherlands and Luxembourg to trade indirect goods and services. Neef (2001, p. 25) defines such indirect goods as ‘any commodity or service that a company buys that does not result directly in finished goods’. Tradcom’s suppliers tend to enjoy a long term relationship with Tradcom. Many of the suppliers now co-own Tradcom. Computer systems between suppliers and Tradcom are highly integrated. Customer data couplings are looser. Since this trade concerns indirect goods, customers typically do not want to invest heavily in linking their systems directly to the platform.

After presenting eight problems associated with sharing data in Section 2, Section 3 illustrates how the problems take a different form in different data sharing configurations. This leads to the conclusion that those problems have to be considered when evaluating inter-organizational data integration configurations. The

Reference this paper as:
Goethals, F. G. “Important Issues for Evaluating Inter-Organizational Data Integration Configurations.” *The Electronic Journal Information Systems Evaluation* Volume 11 Issue 3 2008, pp. 185 - 196, available online at www.ejise.com

goal of the paper is not to give an exhaustive list of issues that have to be considered. Rather, eight problems are discussed which were found to be of relevance in several cases.

It should be noted that this paper investigates only the exchange of explicit knowledge. Tacit knowledge is at least until some point in time not stored in computer systems (see Nonaka 1994). As such, there are a number of distinct characteristics which make it inappropriate to treat explicit and tacit knowledge as one and the same.

2. Problems associated with sharing data

Case studies (such as the Tradcom case study) and literature (Ljung 2004; Sheu et al. 2001; Tang et al. 2006; etcetera) indicate that a number of important problems become evident when companies begin to share information. This section identifies eight frequently encountered problems. Each problem is defined and its relevance is illustrated by referring to one or more of the practices presented above: VMI, PLM and e-procurement via Tradcom. This paper focuses on identifying principal problems. Further research is required to elaborate rectification projects.

The eight problems we discuss are the following:

1. First and foremost, companies have to define what information flows are valuable from a business viewpoint: what data does a company want to use, when should it get that data for the data to be useful, etc. Identifying what information flows to automate is not an easy task.
2. When one wants to identify information flows, one will stumble across another problem: if companies want to share data about an object, they have to acknowledge they may use the object in different tasks and they, therefore, may have another viewpoint on the object. Thus they need to map their viewpoints before they can actually go about sharing data.
3. At this stage partners know what data they intend to share. The next phase requires defining the data format.
4. To realize the information sharing, investments will need to be made. Partners have to agree on who will bear costs for installing, maintaining and upgrading the systems.
5. Once the investments are made, business continuity can only be ensured if the data is provided by the systems as needed. The systems should thus offer appropriate service levels in terms of availability, response time, etc.
6. Sharing data gives the parties involved new sources of power. Data receivers may forward the data to a third party or may inadequately secure their systems, the data provider may not pay enough attention to data quality, etc. Partners must preserve the value of the data sharing.
7. A party should be designated that can decide what can or has to happen with specific data, and what is prohibited. This requires clarifying who the designated data owner is.
8. All of the problems described above have to be dealt with in the frame of changing relationships. New partners may be added, and partners may be removed from the network.

2.1 Valuable information sharing practices have to be identified

Partners have to identify what information sharing practices add value.

2.1.1 Problem description

Identifying the content, timing and parties involved for an electronic data transmission is far from straightforward. When starting a Business-to-Business integration (B2Bi) effort it is very likely that the existing information flows between the partners have not been ‘architected’ or made explicit (Goethals et al. 2006b). This means many existing information flows are ‘invisible’, complicating their digitalization. Moreover, creativity and coordination are needed to redesign existing information flows and identify valuable new information flows and ways to realize them. A party can be required to capture new or more fine-grained data internally. New data content can be identified that only exists at the level of the Extended Enterprise.

2.1.2 Illustration of relevance

The introduction of Electronic Data Interchange (EDI) enabled companies to digitally transmit existing documents such as purchase orders. This way, the traditional replenishment process was automated. Later on, creative minds introduced VMI as a substitute for the traditional replenishment process, relying upon the introduction of new information flows that were also based on EDI. Sellers only used to receive (and to have knowledge about the reseller's purchase orders. With VMI, sellers regularly receive information about the resellers' sales orders and stock states (Angulo et al. 2004; Danese 2004). Pursuing VMI requires agreement on what messages to transmit and the timing of message transmissions. Messages may only be transmitted after a request or regularly as a subscription with a publication once a week, daily, several times per day, etcetera.

PLM acknowledges regular transmissions of information between different designers are valuable. Sharing information among different designers requires many decisions to be made, the timing of transmission for example. If one is reworking an artifact that is being used by others in their decision making process, should a new version of the design then be transmitted to the others every 15 minutes, every hour, every day or every week?

2.2 Partners have different viewpoints on objects

The determination of what data to share is complicated by the fact that the partners have a different perception of the objects on which they want to share data.

2.2.1 Problem description

Successful data exchanges between two parties generally fit with theories on boundary objects (or trading zones) (Star & Griesemer 1989; Chrisman 1999). Boundary objects are objects that both inhabit several intersecting social worlds and satisfy the informational requirements of each of them. 'Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites' (Star & Griesemer 1989, p 393). A 'flight' is an example of a boundary object:

- A pilot flying an airplane needs to know where a flight is going to, but not who is sitting where on the plane.
- The stewards have to know where everyone is to be seated on the plane.
- The catering company needs to make sure that the right type and quantity of food is available for a specific flight. Appropriate catering is a function of
  - the number of people seated on the plane, what is related to the identity of the people on the plane, known by the stewards; and possibly of
  - the location where the plane is heading to/from, known by the pilots.

Carlile (2002) asserted that knowledge is localized and embedded. Integrating the data about an object is problematic as different user groups have different perceptions of the object, having common and diverse needs and interests. They execute different tasks using the object data and different business rules pertaining to the object are relevant. User groups' views are interrelated and information in different views needs to be aligned.

The ultimate level of transparency is to integrate enterprises so that only one electronic version of an 'order' is saved in a shared database and that Purchase Order and Sales Order melt entirely together. Making this vision reality is difficult because companies involved interpret the concept of 'order' differently. That is, an 'order' is a boundary object. Simply stated, different companies can have a number of common *and* specific events that may affect the state of the object. For common events, it has generally been acknowledged that the lifecycles of the different perceptions of the boundary object should be compatible. Snoeck et al. (2004) for example show that both companies should agree whether an order first has to be paid for and then has to be delivered; or the other way around. This does not create an additional burden on the goal to store only one copy of the boundary object. However, when it comes to the specific events, the boundary object is brought in a different state for one company, but not in another state for the other company. For example, checking the customer history can categorize the order into the state 'urgent handling' or the state 'postponed handling' for the supplier, while the customer leaves it in the state 'PO sent'. Consequently, not all information on the boundary object simply can be stored in a single shared electronic object.

2.2.2 Illustration of relevance

Carlile (2002) studied four principal functions that are dependent upon each other in the creation and production of a product: (1) sales/marketing, (2) design engineering, (3) manufacturing engineering and (4) production. He found that knowledge specialization makes it complex to work across functional boundaries and to accommodate the knowledge created in another practice.

People perceive product data differently. Designers make a Computer Aided Design (CAD) file and machine operators focus on Numerical Control (NC) files. Workmen that install a part are only interested in the outside appearance of the part and whether that part fits. A designer considers the same object as vector graphics. Consumers of a product want to see how the product appears in different settings (e.g., a refrigerator placed in different kitchens). Engineering Bill of Materials (BOM) can be changed to create a manufacturing BOM by adding information about machines that will be used during production. Although all data concerns the same product, only to some extent are the various parties interested in the same information.

To determine the relationship between data one has and data others need, the events that happen to an object have to be investigated. Several parties influence the state of the object and the data that is *valid* is thus determined by many parties together.

All partners need data that is valid to them. This validity does not require consistency, as events happen to the object that are not relevant to a party. Data that is valid for one company is not necessarily valid for another company. For instance, an engineering company decides to forward its entire engineering BOM to a number of suppliers, each of which is only interested in a subpart of the BOM. Updates to the BOM are relevant for one party, but not for another. That is, without the update the data is not consistent, but still valid for some suppliers.

2.3 An appropriate data format has to be defined

Once the information companies want to share is known, partners have to look for a way to share data so that it offers desired ‘functionality’.

2.3.1 Problem description

Choosing a data format is not easy. Partners can have data in systems they have developed themselves or that were created by vendors. These systems are likely to support different data formats and their interfaces may not be documented. Even though many XML industry standards have been developed to exchange data, many companies have legacy investments in EDI systems. These make the choice for XML messages less evident. An important decision companies have to make concerns how they annotate and structure data. Different formats can be adopted to transmit data. Different formats, however, enable different ‘functionality’.

If data is meant to be ‘fully functional’ on another computer system, using the data will be greatly facilitated if the data is annotated. That can be done by sending data in a standard XML format. Hence, the sender can try to standardize and enhance the data structure, thus transforming unstructured data into semi-structured data.

Senders may not like the recipient to use information for unknown or undesired purposes. Hence, they may transmit data that is not fully functional. The receiver can be hampered because the data is not annotated and highly unstructured.

Data that is highly structured at the sender’s site, such as prices of products in a relational database, can be transmitted in unstructured documents such as highly graphical brochures. An information owner who wants to prevent poaching and the like (see Section 2.6) may adopt a format that reduces functionality, thus limiting the value of the transmission for the information receiver (Clemons & Hitt 2004).

2.3.2 Illustration of relevance

The Tradcom case indicates that the data format that companies internally use often matters. This is especially relevant for companies that limit relationships to short term contracts. Tradcom has a short term relationship with the customers. Therefore, it allows the customers to send orders in the format they use internally, such as SAP (Goethals et al. 2007a). This lowers the burden on the customers to do business with Tradcom. However, not every internally used format is accepted by Tradcom. The platform ‘understands’ messages from counterparts in *a number of* vendor formats (such as SAP). Customer’s proprietary system formatted messages are generally not understood (Goethals et al. 2007a). From the supplier-side, where there are long term relationships, Tradcom-specific XML protocols have been developed. All suppliers have enabled their systems to speak that Tradcom-specific language. Their internal data storage format thus is of no direct relevance.

Companies can seek to restrict 'data functionality'. As an example, a supplier may not want potential purchasers to load the entire product catalogue into their ERP systems. This would make it too easy for them to compare different suppliers'prices. Still, it is valuable for the purchaser that data on products he buys from the catalogue are made persistent in his system. A solution to this problem was defined by SAP and Ariba (SAP 2003) and implemented at Tradcom. With the Open Catalog Interface (OCI) or 'punch-out' solution, a purchaser sees the catalogue on a webpage formatted in HTML. The catalogue does not enter the ERP system of the purchaser. The purchaser can select the products he wants to order via a web interface and can have a standardized message sent to his ERP system so that only the data on the desired products is posted to his ERP system.

Many suppliers do not want purchasers to download the entire catalogue into their system because this would make it too easy for the customers to benchmark offerings. This is, however, not always the case. If the competitive strength of the company is just that it *has* the lowest prices, it *will* try to make its price data as accessible as possible. In case the supplier tries to differentiate his products from those of competitors using other variables than the price, the supplier will try to pull the attention of the customers to the appropriate variables; for example by presenting a nice picture on the catalogue-webpage. Short term buyer-supplier relationships tend to work this way. Suppliers in longer term relationships may make prices available in an accessible format, even if these do not show the best side of the supplier.

When considering PLM, for a designer a bit map picture is not very useful: they require vector formats to be able to make changes to the product's design.

2.4 Different parties have to make investments

Investments are needed to enable information sharing. Partners have to agree who will bear the costs for installing, maintaining and upgrading systems.

2.4.1 Problem description

Investments can be made primarily by one company or split among the partners and/or third parties. If one party changes or withdraws the systems he possesses, other parties may have to make further investments in order to ensure connectivity. Without their investments, information flows will fail if interfaces for example have become invalid. Long term relationships partners are better placed to know who currently uses a modified system and are able to negotiate, plan and control changes.

It can be desirable that new entrants to the network do not need to make significant capital investments. If big investments are required, this complicates the decision to commit to the network because they then are locked-in to a relationship that may not end in a partnership. On the other hand, a party that leaves can take a part of the investments with him so that the rest of the Extended Enterprise suffers more than proportionally from the departure of that party. That is, they lose more than just the connection with that party: they also lose connection with other parties.

Han et al. (2004) showed companies tend to be reluctant to share information via a non-neutral infrastructure. The natural fear is that the information will be used beyond a contractual relationship. This seems significant in cases of coopetition (i.e., cooperation between competitors).

It is thus not only important to decide who pays for the investment just because it *negatively* influences who wants to share data (i.e., as it costs money you do *not* want to share data), but also because it has a *positive* influence (i.e., as you own the storage space yourself you *are* willing to share data).

2.4.2 Illustration of relevance

The spread of investments by collaborators depends on the chosen information sharing configuration. Therefore, no general comment on VMI or PLM can be given.

In the Tradcom-case (Muylle & Croon 2003), customers did not have to make big investments: they only had to enable their systems to send purchase orders in their proprietary format to Tradcom. Their suppliers had to make bigger investments: they needed to communicate in a Tradcom-specific XML format with Tradcom.

Tradcom took care of translating between the customer's formats and Tradcom's standardised format. A group of suppliers together own Tradcom and they are thus the parties that invest most in the communication. Customers can easily enter and leave the marketplace whereas suppliers are closely tied to Tradcom. Tradcom reduces investments for individual suppliers because the lowest cost option is to have suppliers connect only to Tradcom, and this interfaces with systems of all group's customers. If a new customer enters the marketplace, the individual suppliers do not have to make new investments. If a new supplier enters the marketplace, other existing suppliers do not have to make investments. Neither do their customers.

Should one or several of the Tradcom-owning suppliers 'leave' Tradcom and take the Tradcom platform with them, the customers and all other suppliers would be disconnected.

2.5 Partners become dependent upon service levels provided by data sharing systems

The value of data sharing systems depends on the service levels that are realized.

2.5.1 Problem description

A significant service level requirement is 'availability'. Lack of data inhibits the functioning of companies. User's service level expectations have to be managed so that unexpected temporary events that cause lower availability do not create negative sentiments (Eliadis 2007).

Other service level issues that require attention are system response times, the variability in response time, the maximum time span specific data is inaccessible, etcetera (Hartley 2005). Current 'embodied' solutions to deal with contingencies, such as calling a colleague in the partnering company, can become less evident once the computer systems are integrated. This is because computer systems integration can weaken the personal relationships that exist between companies (Levina and Vaast 2006). This increases the dependence upon sustained peak performance of the computer systems.

2.5.2 Illustration of relevance

In the context of VMI, unavailability of the reseller's stock and sales data prevents the seller from accurately defining the 'purchase order quantities' (i.e., the 'sales order quantities' from his viewpoint). The network may be able to function without this information, but *only for a limited time span*.

Similarly, companies sharing product data can usually execute their tasks if they have access to not-so-recent product data. Still, sometimes it is important an update *is* considered in the execution of a task. The infrastructure then should allow the timely communication of this update. Product data files can be quite large. Files of more than 100 MB are common. If such a file is 'suddenly' needed, the requesting system may have to wait a long time for the transmission to complete, disrupting the system's execution.

2.6 Partners must preserve the value of data sharing activities

The partners should handle data in line with how other partners would like them to handle it so that they can build trust.

2.6.1 Problem description

There are two main situations that create trust between companies: the receiver has faith in the quality of the data and the sender has faith in the receiver preserving the confidentiality of the data. This trust can be marred by the partner himself, or by a third party. This is shown in Figure 1.

	Sender (S)	**Receiver (R)**
Himself (H)	Quality	Poach
Third party (T)	Corrupt/Steal	Steal

Figure 1: Sender and receiver are both responsible for preserving trust

Receivers have several responsibilities. A party that obtains new data has the power to 'poach'. According to Clemons and Hitt (2004, p94), 'poaching' comprises three components: '(1) the exchange of information between two parties, as a natural by product of contractual exchanges for other goods or services, necessary for the performance of contractual obligation; (2) the subsequent use of this information by the receiving party, outside the purposes for which the information was provided, and for its own benefit or economic gain; and (3) at the expense of, or creating economic damage to, the party that provided the information’. By extension, poaching cannot only concern ‘production data’, but also meta-data: partners can pass on data about what data is being shared.

Going one step further, receiving parties may reject poaching but inadequately secure their systems so that unauthorized parties can gain access to the data. In essence, a network is only as secure as its weakest module or connection. Security requires authentication and authorization (Logan 2004). Authorizations typically are discussed in the context of *who* is trying to access the data content. Goethals et al. (2006a) found authorizations tend to depend on what task the data will be used for, the physical location the message has to be sent to, the medium over which the message is transmitted, the moment the message is sent, the data content, the message format, whether it concerns a single record or a batch of records, whether the data is coarse-grained or fine-grained, etcetera. Authorizations are thus a complex matter.

Both the receiving and sending parties’ systems should be trustworthy. The sending party should make sure that it provides data of the expected quality. Therefore, this party has to prevent stored data corruption and ensure accurate data entries. This can be a significant task for operators that do not know or understand the relevance to users.

A special problem concerning the preservation of the functionality of the data has to do with non-repudiation. Typically, a party executes tasks in response to received messages. Parties that have taken part in information transmissions should not deny their participation later. In a B2B context, the concept of non-repudiation is important. This concept embraces two ideas: the sender cannot deny that he sent the message and the receiver cannot deny that he received the message (Kremer et al. 2002).

‘Privacy’ is an often mentioned topic in the context of partnering companies. Clarke (1999) asserts information privacy requires that information about individuals should generally not be available to others and that, where such data is possessed by another party, the individual should be able to control the data and its use to a considerable extent. This leads to the conclusion that a customer that gives personal information to a specific company, not automatically agrees this information can be forwarded to partners of this company. Partners clearly should handle data with care, fully respecting the agreement the originating organization has with its customer.

As the partner does not have a direct relationship with the customer, chances of poaching are elevated. Poaching can be limited by using a data format with limited functionality (Clemons & Hitt 2004). Logically, one would not expect poaching to show up in an Extended Enterprise context as poaching potentially damages trust. Still, Clemons and Hitt (2004) state actual poaching is difficult to observe, turning it more plausible. They assert poaching is more likely to turn up if there is weak intellectual property protection (e.g., the impossibility to ‘return’ the information at the end of the contract) and if poaching was not prohibited in a contract. These are two conditions that often apply in a close partnership.

One could make a similar assumption about the problem of securing the systems. One could assume that partners ensure the security of their systems because they are dealing with confidential information. In one study, Dynes et al. (2005) indeed found firms do not formulate big security requirements for their suppliers. Yet, in an earlier study Dynes (2004) found that, in general, companies *are* auditing the information security status of potential partners. One difficulty with such assessments is that they are said to slow down partnering.

2.6.2 Illustration of relevance

Applying Figure 1 to the PLM example suggests the following risks:

(RH) The party can consciously share his partner’s designs with a competitor of that partner.

(RT) If a party’s systems are not secured well enough, intruders may gain access and steal the product designs that are owned by this party’s partner.

(SH) For one party it can suffice that data are generally precise, while for another party extreme accuracy may be relevant. Rounding done by the former party (implicitly) results in inaccurate data for the latter party.

(ST) If an external party can hack the systems and can change measurements, the plans become worthless. If such problems re-occur this makes it hard for companies to have confidence in the data that is being shared.

Similar risks exist in the VMI case.

2.7 Data ownership has to be unambiguous

A data owner has to be appointed. A party decides what can or has to happen with the data and what cannot. Often it is not clear who is the designated data owner.

2.7.1 Problem description

On the one hand, *several* parties may want to decide what can or must happen with the data. On the other hand, *no* party may be designated as data owner or may feel responsible.

Data ownership is defined in the California State University Data Warehouse Glossary as the "responsibility for determining the required quality of the data, for establishing security and privacy for the data and determining the availability and performance requirements for the data". Data owners have the authority, accountability and responsibility to create and enforce organizational rules and policies for business data. They have the right to determine access, creation, standardization and modification rights (Alstyne et al. 1994). If ownership is not well defined, data that is considered to be valid may in fact not be valid; may fall in the wrong hands, etcetera.

Levina and Vaast (2006, p28) indicate that, in community-like practices, boundary objects are co-produced and represent negotiated outcomes of a joint practice. If different partners collaboratively change a view on an object, it becomes obscure who actually owns that specific view on the object. As a result, the ownership of intellectual property is one of the most critical legal risks that confront Extended Enterprises (Bader 2008).

2.7.2 Illustration of relevance

If a subpart of a product is created by a partner it is not always clear who is the owner of the subpart-data: the producer of the final product, or the designer of the subpart. Similarly it may be unclear who the owner of the stock data is in case of VMI. Unless consignments are used, the stock is the property of the reseller. This would suggest the reseller is the data owner. The stock is, however, managed by the seller. So, the seller could be seen as being the data owner. The fact that the reseller and the seller have to agree on a stocking plan, inventory turns, replenishment rates, delivery frequency and the like shows that both parties have something to say about the stock. The key issue here is defining responsibility, accountability and authority.

Data ownership problems also show up in cases where there are data that only exist at the level of the entire Extended Enterprise (i.e., at the level of the collection of the collaborating legal entities). In Tradcom's case, suppliers sell their products through Tradcom to a multitude of customers. Although individual suppliers only get orders with respect to their own products, there is information available at Tradcom-level about orders that contain products from different suppliers and about customers that entered the platform to do business with one supplier but in time started doing business with other suppliers. The question is who owns this data and has responsibility for managing its quality. This information can be very valuable yet easily overlooked because it is by nature not really owned by any of the individual parties.

2.8 The involved parties change over time

All of the problems above have to be dealt with in the frame of changing relationships. Partners can be added to and can be removed from the network.

2.8.1 Problem description

The parties that *provide* data and the parties that *use* data change over time. Data that is *valid* changes if partners change. Also, there are changes in the agreement on which parties can access data: some parties can *probably* no longer access the data.

Partners that are dropped from the Extended Enterprise are likely to become competitors of the Extended Enterprise. Retained access to the Extended Enterprise data may hamper the competitive position of the Extended Enterprise. However, it can be desirable that they still get access to specific data, for example to offer after-sales service. Similarly, it is often important for members of the Extended Enterprise to maintain access to former partner's data.

2.8.2 Illustration of relevance

When a supplier of a semi-finished product is replaced, the valid product data set typically also changes. That is, in the design of the assembled product, the design of the semi-finished product has to be replaced by the design of comparable items made by the new provider, or at least its interface. While the old design data may not be valid for newly produced products it remains useful to maintain old products and the after-sales service provision. This implies that different versions of the component design have to be managed and related to specific variants of the produced product.

In general, the former supplier of the component no longer should have access to the other product data as they could disperse this data across the new network they have entered. Nevertheless, if this supplier still is to create spare-parts for the product under consideration, they shall need to be given access to specific parts of the data warehouse.

With VMI the seller truly obtains insight in the reseller’s business (Monitor Technologies 2005). This can be illustrated with the fact that if a transmission of sales and stock data fails, the seller, in practice, still has quite a good idea of the sales and stock levels at the reseller’s site because he knows his business so well. This fact is very desirable while both companies do business with each other. However, once their relationship is suspended, the seller could abuse the knowledge gained from this relationship.

3. Discussion: the problems are important when evaluating data sharing configurations

The eight problems presented above should be taken into consideration when deciding upon an inter-organizational data integration configuration. Many configurations exist that have radically different properties in terms of how the eight problems show up.

According to Goethals et al. (2007b), one can distinguish between different data integration configurations on the basis of

1. how decisions with respect to the data sharing are made: decisions with respect to data sharing can be taken centrally or decentrally. A central consortium such as RosettaNet can define a global data model, what messages can be exchanged, etcetra. Alternatively, parties can bilaterally agree on the content that will be shared, on an XML Schema they shall use, etcetera.
2. how data is transmitted: Data can be transmitted directly from data provider to data receiver or data transmissions can happen via a central system. Such a central system then can take care of format mappings, guaranteed delivery and other services.
3. how data is stored: The data that is to be shared may be stored in one central repository that is shared by all partners, or can be stored in different locations of different partners.

A data integration configuration is characterized by the extent to which it is rather centralized or rather decentralized on these three dimensions. Depending on the degree of (de)centralization, a configuration has different properties in terms of the problems presented in Section 2. Each configuration has good points and bad points to note for each problem.

In CIMdata (2005), an **entirely centralized data integration configuration** is presented where companies create a single shared data repository that contains all the data the companies want to share. This configuration can be evaluated in terms of the problems presented above. For example, it was stated that timing, content and responsibilities for transmissions as well as data access rights have to be determined. When deciding upon necessary updates of the data in the shared storage space and access rights, a completely centralized configuration has the disadvantage that an update to the central data is relevant for one party, but may need to be hidden for another party. Also, for all information requirements that should be fulfilled there is this third party, the shared storage space, which needs to be involved. This central party can have established policies that are considered rigid and inappropriate by local offices. This may reduce motivation to introduce new information flows (Streeter 1973). However, the fact that there is an overall view available on what information sharing is happening makes it easier to manage the whole system.

As another example, when it comes to the data format, the intermediary can take care of a big number of tasks, among which the transformation of message formats and the transformation of content. The product number used in one system, for example, can be reformatted to suit product numbers in another system. However, adapting the ‘functionality’ of the data may be a difficult matter in this setting. That is so, not only because it is harder for a central office to understand local functionality requirements and restrictions, but also because the data can be available in the *commonly owned* shared space in a fully functional form. Furthermore, when considering service levels, there is a single point of failure (Iamnitchi 2000). All parties use the capacity of that same central point and the capacity that is available for one party thus depends upon the capacity used by all other parties. Still, as many parties co-invest in central systems, they may reasonably expect a high-quality shared space, operated by specialists who know how to intervene in case of problems. As a final example: as there is only one site with data, there is only one site that needs to be protected from security breaks (Yu 2000). However, if this security system fails a non-trusted party can immediately gain access to all data.

Centralized configurations can be contrasted with **entirely decentralized data integration configurations**, where partners set up direct data exchanges between the system of the partner that wants to use the data and the system of the other partner where the data is available. In this case, parties autonomously can decide what data they want to share and are less concerned with high-level restrictions. Mintzberg, (1992) considered this to be a stimulus for innovation. However, setting up many point-to-point connections complicates management of the entire integration effort. Also, parties do not share the capacity of a central system. This lowers the chance that another party is using some capacity one had planned for. This, however, eliminates the possibility to use capacity that is currently not used by other parties. Capacity can thus be available in the wrong place (HP 2003). Furthermore, implementing security policies in a highly heterogeneous environment is harder (Iamnitchi 2000). Several data stores and transmission systems have to be secured and it is unclear what the security-level is of the Extended Enterprise as a whole and it is hard to identify whether anyone is poaching. There is no central intermediary that can oversee the unusual behavior of a party that are breaking trust.

The illustration above highlights that companies are confronted in different ways with the problems from Section 2 if they choose another data integration configuration. Consequently companies have to find out how important a specific problem is to them and consider the identified problems when evaluating different configurations.

Many configurations exist (Goethals et al. 2007b). The goal of this paper is not to give an evaluation of all integration configurations. This paper discusses problems that repeatedly have shown up when such configurations were installed and identified that these problems take on different forms when different configurations are used. Further research is needed to evaluate different configurations with respect to the problems discussed in Section 2. Management is advised to consider problems in the context of a specific configuration they have in mind.

4. Conclusions

This paper discusses eight problems that show up when partnering companies decide to share information. Each problem occurs in different practical situations, as illustrated with examples from the PLM and the VMI domains and the Tradcom B2B trading platform.

It is important to deal with each of these problems. It was shown that the problems take a different form in different Business-to-Business data integration configurations. Companies that need to choose an inter-organizational data integration configuration should determine appropriate weights for the problems.

One of the problems concerns the distribution of investments and benefits. The amount of counterparty-specific investments companies are willing to make depends upon the expected duration of the use of the investment and thus upon the expected duration of the relationship between the companies. Furthermore, a data-owner is needed at inter-organizational level which determines who can access what data and those that may not. The value created from using specific information sharing configurations depends on service levels that are realized and the way partners handle data. Senders are dependent upon their partners to reject poaching and to secure their systems. Receivers are contingent upon the senders for transmitting the data in an appropriate quality. Data format problems and obscurity about appropriate data flows further complicate the issue.

Acknowledgements

I gratefully acknowledge the help from Prof. David Newlands in writing this paper.

References

Alstyne, M. Van, Byrnjolfsson, E. and Madnick, S.E. (1994) "Why not one big database? Principles for data ownership", *MIT Sloan School Of Management Working Paper Series* 3695-94, p 39.

Angulo, A., Nachtmann, H. and Waller, M. (2004) "Supply Chain Information Sharing in a Vendor Managed Inventory Partnership", *Journal of Business Logistics*, Vol. 25, No. 1, pp101-120.

Bader, M.A. (2008) "Managing intellectual property in inter-firm R&D collaborations in knowledge-intensive industries", *International Journal of Technology Management,* Vol. 41, No. 3/4, pp311-335.

Carlile, P.R. (2002) "A pragmatic view of knowledge and boundaries: boundary objects in New Product Development", *Organization Science*, Vol. 13, No. 4, pp442-455.

Chrisman, N. (1999) "Trading Zones or Boundary Objects: Understanding Incomplete Translations of Technical Expertise", *Proceedings of the Society for Social Studies of Science Annual Meeeting*, San Diego, October 28-31, p15.

CIMData (2005) "Eurostep's Share-A-space – Making your intellectual assets the backbone of your information management strategy". [online], Eurostep, http://www.eurostep.com/files/Share-A-space_CIMdata_2005.pdf.

Clarke, R. (1999) "Internet Privacy Concern Confirm the Case for Intervention", *Communications of the ACM*, 42, 2, (February 1999) pp60-67.

Clemons, E.K. and Hitt, L.M., (2004) "Poaching and the misappropriation of information: transaction risks of information exchange", *Journal of Management Information Systems*, Vol. 21, No. 2, pp87-107.

Danese, P. (2004) "Beyond Vendor Managed Inventory: the Glaxosmithkline Case", *Supply Chain Forum: An International Journal*, Vol. 5, No. 2, pp32-40.

Dynes, S. (2004) "Security and Privacy: at odds with speed and collaboration?", *Executive Round Table Series*, [online], Tuck School of Business, http://mba.tuck.dartmouth.edu/digital/Programs/CorporateEvents/SecurityAndPrivacy/Overview2.pdf p. 13.

Dynes, S., Brechbühl, H. and Johnson, M.E. (2005) "Information security in the extended enterprise", *Proceedings of the 4th Workshop on the Economics of Information Security*, Cambridge, MA.

Eliadis, H. and Rand, A. (2007) "Setting expectations in SaaS: The importance of the service level agreement to SaaS providers and consumers", [online], Software & Information Industry Association, http://www.siia.net/Software/pubs/saas_0207.pdf, p. 14.

Goethals, F. G., Vandenbulcke, J., Lemahieu, W., Snoeck, M. and Cumps B. (2005) "Two Basic Types of Business-to-Business integration", *International Journal of E-Business* Research, Vol.1, No. 1, pp1-15.

Goethals F. G., Lemahieu W., Snoeck M. and Vandenbulcke J. (2006a) "Dependencies between Data Decisions in the Software Process", *Proceedings of the SPW/ProSim Workshop*, Shanghai, May 20-21, 2006, LNCS Volume 3966, pp48-55.

Goethals, F., Lemahieu, W., Snoeck, M. and Vandenbulcke, J. (2006b) "Management and Enterprise Architecture Click", Information Systems Frontiers, Vol. 8, pp67-79.

Goethals, F. G., Lemahieu, W., Snoeck, M. and Vandenbulcke, J. A. (2007a) "The data building blocks of the enterprise architect", *Future Generation Computer Systems*. Vol. 23, No. 2, pp269-274.

Goethals, F. G., Snoeck, M. and Lemahieu, W. (2007b) "The Business-to-Business Integration solution space", *Pre-ICIS Workshop on E-Business*, Montreal-Canada, December 9, 2007.

Han, K., Kauffman, R.J. and Nault, B.R. (2004) "Information exploitation and interorganizational system ownership", *Journal of Management Information Systems*, fall 2004, Vol. 21. No. 2, pp109-135.

Hartley, K.L. (2005) "Defining effective service level agreements for network operation and maintenance", *Bell labs technical journal*, Vol. 9, No. 4, pp139-143.

HP Invent (2003) "Architectures for storage consolidation", [online], HP, http://h30046.www3.hp.com/uploads/whitepapers/Architectures_for_Storage_Consolidation_whitepaperF.pdf.

Iamnitchi, A. (2000) "Centralized vs. Decentralized design in Internet-Scale Applications", *Proceedings of the International Conference on Parallel Processing*.

Kremer, S., Markowitch, O. and Zhou, J. (2002) "An intensive survey of fair non-repudiation protocols", *Computer Communications*, Vol. 25, No. 17, pp1606-1621.

Levina, N. and Vaast, E. (2006) "Turning a community into a market: a practice perspective on information technology use in boundary spanning", *Journal of Management Information Systems*. Vol. 22, No. 4, pp13-37.

Ljung, R., Johansson, M. and Rosén, J. (2004) "Sharing product data through life in the Extended Enterprise", *Product Data Journal*, Vol. 11, No. 2, pp40-42.

Logan, C. (2004) "Security principles: identity, authentication, authorization", *Current IS&T Issue*, Vol. 20, No. 2. pp 5.

Mintzberg, H. (1992) *Structure in fives*, Prentice Hall, p 312.

Monitor Technologies (2005) "Remote and Vendor Managed Inventory Management Systems for Bulk Solids", [online], Monitor Technologies, www.monitortech.com/webseminar/RMI_V2doc.pdf

Muylle, S. and Croon, E. (2003) "Venturing into e-commerce in the MRO market", ECCH, p 19.

Nonaka, I. (1994) "A dynamic theory of organizational knowledge creation", *Organization Science*, Vol. 5, No. 1, pp14-37.

Neef, D. (2001) *E-procurement*, Prentice Hall Upper Saddle River, p 1996.

SAP (2003) "Open Catalog Interface (OCI) Release 4.0", [online], SAP, https://www.sdn.sap.com/irj/servlet/prt/portal/prtroot/com.sap.km.cm.docs/library/icc/Open%20Catalog%20Interface.pdf.

Sheu, R.K., Chang, Y.S. and Yuan, S.M. (2001) "Managing and sharing collaborative files through WWW", *Future Generation Computer Systems*, Vol.17, No.8, pp1039-1049.

Smaros, J. Lehtonen, J.M., Appelqvist, P. and Holmström, J. (2003) “The impact of increasing demand visibility on production and inventory control efficiency”, *International journal of physical distribution & logistics management*. Vol. 33, No. 4, pp336-354.

Snoeck, M., Lemahieu, W., Goethals, F., Dedene, G. and Vandenbulcke, J. (2004) “Events as atomic contracts for application integration”, *Data and Knowledge Engineering* 2004 Oct; Vol. 51, No. 1, pp81-107.

Star, S.L. and Griesemer, J.R. (1989) “Institutional Ecology, Translations and Boundary objects: amateurs and professionals in Berkeley’s museum of vertebrate zoology, 1907-39”, *Social Studies of Science*, Vol. 19, pp387-420.

Streeter, D.N. (1973) “Centralization or dispersion of computing facilities”, *IBM Systems Journal*, No. 3, pp283-301.

Takalikar, S. (2004) “Leveraging PLM for manufacturing excellence”, *Express computer,* April 2004, [online], expresscomputeronline http://www.expresscomputeronline.com.

Tang, M., Lee, B.S., Tang, X and Yeo, C.K (2006) “The impact of data replication on job scheduling performance in the Data Grid”, *Future Generation Computer Systems*, Vol. 22, No. 3, pp254-268.

Yu, E. (2000) Centralize or decentralize? “A requirements engineering perspective on Internet-scale architectures”, *Proceedings of the International Conference on Parallel Processing*.