Improving the Quality of Information Exchange in a Business Chain

A.T.M. Aerts[1], M.J.H. Pustjens[1], A.F.L. Veth[2]

[1] *Department of Mathematics and Computing Science,*
Eindhoven University of Technology,
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
Email: wsinatma@win.tue.nl

[2] *Department of Industrial Engineering,*
Eindhoven University of Technology,
P.O. Box 513, 5600 MB Eindhoven, The Netherlands,
and
ACF Holding N.V.,
P.O. Box 5, 3600 AA Maarssen, The Netherlands
Email: tveth@brocacef.nl

Abstract

Despite the usage of a consistent set of EDI-messages and protocols, EDI-projects too frequenty don't produce the benefits that are expected from them. An important reason for this are discrepancies in the reference data used by the various parties in the information exchange process. In this paper we report on a project in a business chain where this problem has been tackled and discuss the solution that has been proposed and implemented.

1 Introduction.

Many companies are using EDI nowadays to improve the exchange of information needed to support their business relations with other companies[1]. EDI is defined here as the electronic interchange of normalized and standardized data (messages) between the (automated) information systems of different companies [2]. EDI therefore is a suitable means for supporting collaboration between independent parties, especially where routine communication is concerned. Usage of EDI at least has the promise of offering a number of well known advantages over, e.g., communication via telephone, fax or regular mail, such as reduction of administrative costs and data entry errors, decreased order lead times, and improved business relations [3]. EDI can be regarded as a way of implementing the communication between (database) applications of different, independent parties.

The emphasis in many EDI-projects so far has been placed on setting up and implementing a system of standardized messages and of protocols to control the exchange of these messages [4]. Much effort is spent on deciding the structure of the messages, i.e., which type of information should be a part of a particular type of message. Also, a lot of effort is spent to fix the interpretation of the arrival of a particular type of message at a certain stage of the communication process and the proper ways to respond to it. The allowed sequences of messages in an exchange of messages are laid down in the exchange protocols. All this is, of course, quite necessary to formalize the communication patterns, such that they may be handled by automated information systems. A full EDI-implementation should operate in the "unattended" mode of the system: no intervention of a human operator at the level of message exchange should be needed. Communication between two parties using EDI should be at the application level. Only in this way reductions in order lead times and such may be achieved.

Deployment of a consistent set of messages and protocols is sufficient to support communication at the conceptual (intentional) level, but it is not enough. Another necessary condition for realizing the benefits expected from using EDI is that all parties involved speak the same language. It is not enough to use the same grammar; one also has to use the same dictionary. That is, various parties doing business via EDI should refer to the same objects, such as products and services in the same way. Communication should also be consistent at the instance (extensional) level. This is well known, of course, but the issue gets pushed back into the background by the eagerness of parties to start exchanging, e.g., order-messages as quickly as possible.

For example, in a business chain a trading company may buy products from suppliers and sell them to customers. It is very important that every party involved, supplier, trader or customer, uses product references in the order-messages that are understood by the other parties. This is complicated by the fact that a product reference changes, as, e.g., the packaging of a product changes. A trading company will sell a drug by the box to a drugstore that sells it in packages of 10 or 20 pills, but the trading company will order it in bulk from the pharmaceutical company that produces it. The various packagings have distinguishing references. This implies that relations between various references exist and should be clear to be able, e.g., to trace goods through a chain.

This implies that the sets of data stored on products by various parties may overlap and the intersections should be mutually consistent. Data such as the product data in the example above are called reference data, or sometimes state-independent data. The existence of this type of data is independent of the fact whether the products are actually being ordered or delivered or not. It is independent of the momentary state of affairs. Reference data should however be present everywhere in the chain because these data form the basis of the communication between the various parties.

The consistency of reference data is an issue that is often played down or even ignored when EDI is introduced. It surfaces, as soon as the technical infrastucture (hardware, system software, EDI-utilities) has been installed and messages can be exchanged without problems, and the content of the messages becomes important.

Since customer and supplier usually independently manage their own set of product data consistency between these sets is by no means guaranteed. Even when they would start with the same set of data, inconsistency may arise after a while because of differences in the way updates are handled. This is essentially a data management problem, that has to be solved in an environment consisting of semi- and fully autonomous parties, that collaborate in a business chain.

In this paper the results of a project are presented, that was carried out in an environment where product references change frequently, due to changes in product composition and packaging. Particularly in such a situation special care is needed to keep product information synchronized. In the next section we will discuss the background of the problem. Then follows a discussion of the problem and its solution. Finally we will discuss some preliminary findings of the project.

2 Background.

The project took place at ACF Holding N.V., a Dutch company, specializing in the production, marketing and distribution of health care products and services. These products range from consumer products for drug stores, to prescription drugs for pharmacies to the design and installation of a surgical theatre.

ACF comprises four divisions, which together have a total of 18 business units, among which production units and trading companies. From a logistical point of view, the business units of ACF form a distribution chain stretching from producer to customer. Some units act as supplier to others, which in turn act as consumer. The business units collaborate to bring health care products and services from producer to customer. Product data are present at all nodes in this chain. The holding company is concerned with optimising the business performance of the company as a whole. The company operates in a market in which two similar companies are active.

The project decribed in this paper is the second stage of the PRICE-project at ACF. PRICE stands for PRoject Inter Company EDI [5]. The purpose of the PRICE-project is to establish the conduction by means of EDI of business transactions between the business units in order to reduce costs and speed up order processing. At a later stage business units will also start conducting transactions with external partners by EDI. In the first stage of the PRICE-project the technical EDI infrastructure was set up. Communication by means of EDI was implemented for the ordering of products between the production and trading units.

The second phase of the project started in the summer of 1994 and is being concluded with a pilot project. It builds on the infrastructure realized in the first phase. It aims at solving a number of problems that came forward in the evaluation of the first phase.

3 Consistency of reference data: the problem.

A standard scenario [3] for introducing EDI between business partners includes steps such as: identify a promising application domain, set up a pilot project to implement EDI for that application domain, evaluate the results of the pilot project, and if positive, fully implement EDI. The next step after this is to try and identify other candidate applications and broaden the EDI implementation. Usually the implementation of EDI takes the form of introducing one small set of message types and corresponding protocols, and enlarging the number of message types in successive stages.

One of the outcomes of the first stage of the PRICE-project was that the success of using EDI also depends on the consistency of the reference data used by the various parties. In the case of ACF, each business unit involved in the transactions was independently managing its own set of reference data with the obvious consequences for the overall consistency of the data.

The reference data used in the case of ACF are mainly concerned with product data. In the health care market of the order of 150.000 different products are registered, about a third of which are actually in current supply. Of the other two thirds only descriptive documentation is available. The data about these products change in a number of ways. New products are added at a rate of 6.000 or 7.000 a year. This is to a large part due to products for drug stores. Prices of products may change or the packaging (due to a face lift of the product line or special offers!). Legislation may require different or additional data to be stored or distributed with a product. Producers may use ingredients from different suppliers causing the product code to change. Erroneous data may be detected.

It is clear that special care is needed to keep product information up to date and consistent between business units. Without such care, inconsistency of, e.g., product codes will generate exceptions in the automated processing of product orders that have to be dealt with manually. As the volume of orders tends to increase with the usage of EDI, it only takes inconsistency in a small fraction of the total number of orders to undo the advantages of using EDI.

For example, when data in an order message, such as a product code in an order line segment, is not recognized, the message has to be set aside and analysed by a human. Determination of the cause of the error and correction of it, on average takes the order of 15 minutes. One has to contact the sender of the message, e.g. by telephone, one has to determine which product is meant, what the correct code is, and then to make sure the databases of both parties agree on that to avoid future errors. Assuming a rate of 150.000 messages a year (this is of the same order of magnitude as the number of different products in the health care market), and an error rate of 5% this would fully occupy one person. It is clear that the consistency problem has to be fixed before the scope of the EDI implementation can be broadened.

Product data may come from a number of sources. Suppliers may offer new products or change the composition or packing of a product. These product data updates are at present communicated via forms (not standardized) or electronically via diskettes. There is a branche organisation, KNMP, where producers or importers have to register every product and obtain a unique label (code) for it. KNMP publishes every month a tape, containing a survey of all products, currently available. KNMP matches its code with the HIBC code. This way it takes about 6 weeks to publish the information about a new (version of a) product. A third source is the KOMBI-ROM, a CD containing amongst other things, bulk files, used for loading the computers at pharmacies and computers used for ordering at other locations with product data. This CD is also publised monthly. The first two sources are used for updates. The third source, to which the company itself contributes, is used for quality control. Discrepancies with data provided by the competion are looked for and resolved.

Updates of the product databases are done partly automatically, partly manually. This requires a lot of effort, which moreover is duplicated to a fair degree between various business units. For a new product typically 45 fields need to be given a value. Depending on the size of the change, updating information about a single product takes from 2 to 20 minutes. In addition, product data are becoming more voluminous, since the companies are required to provide more information, such as product descriptions ("wrappers") to go with the consumer products.

Discrepancies between data of different business units may arise from various sources. First of all, different business units may need different information about a product. Therefore, some information may be present, some not. Differences in interpretation of the information exchanged should have been noted at the time of the design of the messages and protocols, and have been settled in the mapping of the messages to the local database of a business unit. This is basically a database integration problem. Nevertheless, some differences may have been left unnoticed. Updates may be dealt with differently by various units and also depend on the source that generates it. For instance, the speed of processing updates may depend on the capacity available, such that one unit may complete the updating process well before another one. Also, product information may come directly from the supplier in which case the information may be much sooner available than in case of the KNMP-tape, which takes roughly 6 weeks to prepare and distribute, and then up to 4 weeks to process.

The problem at hand has, as we see, important timing and quality (correctness) dimensions. It is more easily tackled inside an organisation which consists of a number of business units, for which certain sets of data need to be made and kept consistent, than it is between different, independent organisations. However, the solution presented here may also be applied to the latter situation.

4 Consistency of reference data: a solution.

The problem found in the previous section resembles the problems encountered in the management of replica (identical copies) in a distributed database system [6], and centres from a database point of view around synchronizing the updates of the replica. In a distributed database system the management of replica is handeled by a component in the database management system.

The problem in this case has been looked at from a logistical point of view [7]. Information is seen as just another product, at same the level as other products in the business chain, which has to be distributed, subject to time and quality constraints. What needs to be solved then is equivalent to a distribution problem in logistics. An important concept in distribution logistics is the so-called decoupling point, exemplified by the warehouse and the distribution centre, where goods from various suppliers are collected and rearranged to be sent to customers, who ordered them, in such a way that time constraints with respect to delivery are respected at the lowest cost.

The option chosen here is to collect the updates from various sources and apply them to a single database with product information. From this database, the primary or master copy [6], the updates are propagated to those parties in the system to which they are of interest. There are some important differences with the problems in distributed database systems.

The first difference has to do with consistency. For every product a single source has to be designated, that is the owner of the data concerning a product and that is the only party authorized to modify them. This data not only includes characteristics of the product, but also periods of validity of a given set of product data. Typically the producer, or its designated representative, is the one to supply these data. One has to exclude the possibility of multiple sources for the same datum, because this could give rise to mutually inconsistent updates. An unattended (automated) system is not capable of resolving such inconsistencies.

Another important issue is to guarantee users of the system a consistent view. For this purpose updates of product information have to be announced in advance.

The information system has to be able to deal with multiple versions, each with its own validity period. Multiple versions are necessary to prevent the updates to interfere with the ongoing activities in the chain. Product reference updates should not invalidate references in orders that are under way. The time it takes to conclude an update of product data therefore cannot be made arbitrarily small, but still can be made much shorter than the six to ten weeks it takes at present. Also a history has to be kept to prevent errors from orders, e.g., from outside the warehouse system, using expired references.

Another difference has to do with the sharing of the information. Not every customer will want to receive the total of the data brought in, but only a selective portion of it. The system therefore has to offer a mechanism to match supply and demand. In this case a selection mechanism has been chosen, in which the information consumer indicates which fragment of the total data should be in its replica.

A further issue has to do with the choice of data to be brought into the system. Those product data are made available that are of interest to most parties involved. These include the essential product characteristics, and some logistic information, such as producer, distributor, volume, weight, and stock condition. Not included should be internal information, such as production data, and transaction dependent information, such as price and delivery conditions. In addition to product data also another kind of reference data, viz. address data are kept in the system.

The database system is organized as a warehouse storing information. This means, that information is put there by the information suppliers and is distributed from there without modification to the selected information consumers. The information in the warehouse is not used for any other purposes. The ownership of the information remains with the one who supplied it. One central database is used, just for storing and distributing the information. Maintenance of the database is done decentralised, by the information suppliers.

The concept of an Information Warehouse described here differs from the one proposed by IBM [8], which evolved from the Information Repository, used to store metadata, describing an organisation's data and data processing resources. The two concepts have in common that they aim at providing a framework for sharing data among heterogeneous software and computer systems, based on the existence of a shared, central repository. IBM's solution is a corporate one, while the system discussed here is meant to share information between companies.

A difference with a warehouse for distribution of physical goods, is that information, in electronic form, can be duplicated ad libitum. However, the information warehouse has been set up such that only those consumers having selected a product will receive the information.

4.1 Architecture of the Information Warehouse system.

In the previous section it has become apparent that three types of parties are involved in the warehouse system: the information suppliers, the central database (the actual warehouse) and the information consumers. This is depicted in figure 1, where the arrows just indicate the flow of product data.

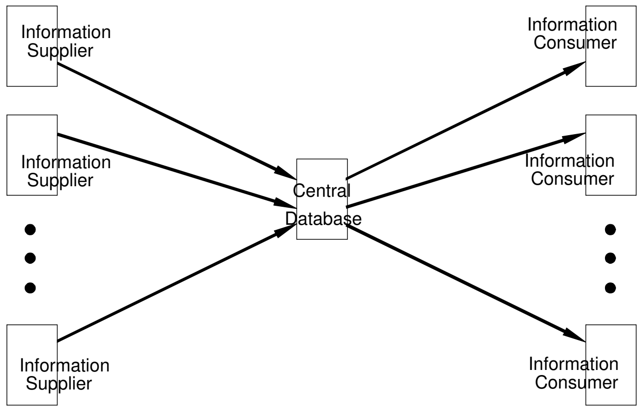

Figure 1: Warehouse Architecture

The Information Warehouse therefore consists of three subsystems with the following functionality:

The **Information Supplier subsystem** allows an information supplier to:

- make updates available to the data in the central database, concerning the products of the particular supplier,
- maintain the address information in the central database of this particular supplier,
- automatically receive the address data of a selected set of users of the central database
- indicate the target group for a specific product and specify also which information consumers are to be excluded from access to this information.

The **Information Consumer subsystem** allows an information consumer to:

- automatically receive updates of information about a selected range of products.
- maintain the address information in the central database of this particular consumer,

- automatically receive the address data of a selected set of users of the central database,
- order information and updates on products. A consumer thus is able to modify the range of selected products.

The **Warehouse (Central Database) subsystem** allows for the:

- storage of data,
- management of users of the central database
- restoration of data in case of disaster
- backup facilities
- automatic updating of product data in the central database on the basis of the updates from the information suppliers
- automatic distribution of updates to selected information consumers
- automatic updating and distribution of updates concerning adresses
- monitoring facilities
- maintenance of the authorizations of information suppliers concerning the product data
- maintenance of the selections of the information consumers

To support the updating and distribution of product information, an EDI message is used.

Every supplier of a product or it's designated representative may act as an information supplier. To become authorized as such a contract has to be drawn up between the supplier and the party responsible for the Warehouse. The supplier then receives the necessary software (supplier subsystem).

Also information consumers have to register with the warehouse. They receive the software needed to maintain their product selections. A consumer selecting a new product receives all data in the database concerning this product and then all subsequent updates. Consumers become aware of interesting products independently of the warehouse activities, via announcements through mailings, advertisements, catalogues and such. The warehouse system clearly does take over all communication activities. This is partially indicated in figure 2.

The main functions of the Information Warehouse are:

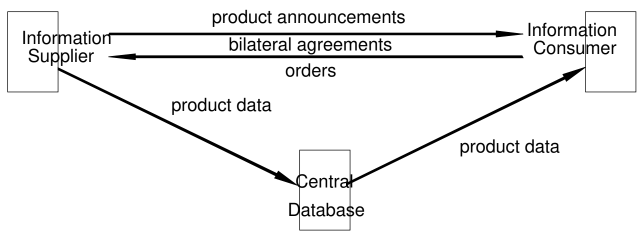

Figure 2: Some of the information flows between various parties

- filter for a particular consumer from the total set of data those data selected by that consumer,
- duplicate and distribute the information concerning a product to alle consumers who selected that product. An information supplier has to provide this information therefore only once.
- act as a decoupling point between information suppliers and consumers. The decoupling takes place with regard to the timeliness of the information exchange and with respect to the means of transport.

The Warehouse provides for a facility which speeds up the propagation of updates. These can be distributed as soon as they become available, in the time scale of hours or a day, instead of weeks. Also all updates come in the same form, such that no special effort is needed to reconcile various formats. There is also a notion of economy built in, since only those consumers interested in a particular update will receive it.

5 Discussion and Outlook.

The goal ACF wants to achieve with the Information Warehouse project is an increase in effectiveness and efficiency of its internal distribution chain in general and its EDI support for it in particular. It is therefore important that as much suppliers and consumers will participate in the project.

Since the product data of ACF cover the complete range of products distributed in the branche the scope of the information in the warehouse extends beyond the company boundary. The warehouse is also of interest to the other parties in the branche. Since these are competitors of ACF, the maintenance of the warehouse is given to a neutral, third party, to lower the threshold for other parties in the branche to participate and so reduce the operation costs.

Within ACF the introduction of the Information Warehouse will lead to a greater efficiency with respect to the management of product data because a lot of duplication of effort in data management is avoided.

Also the performance of the distribution chain is enhanced. Enhancement of this performance is aimed at realizing savings by reducing product stocks. This is achieved by reduction of delivery times, order quantities and order times. These reductions also diminish the room avaible for compensating for errors. A reliable, fast, complete and correct information exchange is needed to achieve this. The product database is at the core of this proces, since almost any communication refers to products. The quality in terms of consistency and actuality of the product database therefore has to be high.

Note that not all data will be multicasted via the warehouse. Bilateral data such as prize agreements, orders and delivery schedules will be communicated directly between suppliers and consumers. Also product announcements will continue to be distributed directly (see fig.2). Only multilateral data are of interest for distribution via the warehouse.

A pilot project is being carried out with 7 suppliers and 5 consumers. Preliminary results indicate a positive effect of the information warehouse on the effectiveness and efficiency of the business processes it supports so far. These result are only qualitative at present.

Some comments can already be made. Product databases tend to contain information on much more products than are at any time being ordered. Only errors (in value or timing) of products being referenced will be disruptive. However, of many products it is not clear whether they are obsolete or just not in demand at the present time. To establish the difference takes a lot of effort. This phenomenon makes it hard to quantify, in absolute terms, the reduction in errors realized by the introduction of the warehouse within the timespan reserved for the pilot.

Another comment is that the warehouse system is not likely to encompass in the near future all parties in the market. Therefore, the system will coexist with the old system of communicating updates. This implies that there will be a choice and an information supplier may at some point prefer a direct communication with a consumer to an update via the warehouse. Think for instance of a situation in which a supplier's stock for some product runs out before an advertised replacement is scheduled to take place. A situation with multiple distribution channels, even more than one warehouse, may be the future.

6 Acknowledgements

It is a pleasure to thank H. Kappert of ACF Holding N.V. and G. Veurink of Brocacef N.V. for many stimulating discussions.

References

[1] V. Leyland, "Electronic Data Interchange", Prentice Hall International, 1993.

[2] W.J. Hofman, "A conceptual model of a Business Transaction Management System", Tutein Nolthenius, Amsterdam, 1994.

[3] P. van der Vlist et al. "EDI in trade", Samsom Bedrijfsinformatie, Alphen aan den Rijn, 1992 (in Dutch).

[4] A.T.M. Aerts, W. Hofman, L.J. Somers, "Message Flow Control", in: Proceedings of the IFIP WG8.1 Working Conference on Information System Development for Decentralised Organisations, Trondheim, 1995.

[5] M.R.W.M. Sonneville, "Intercompany EDI: Implementation of EDI between business units of ACF Holding NV", Master Thesis, Department of Industrial Engineering, Eindhoven Technical University, 1994 (in Dutch).

[6] M. Tamer Özsu, P. Valduriez, "Principles of Distributed Database Systems" Prentice Hall International Editions, 1991.

[7] R.H. Ballou, "Basic Business Logistics: Transportation, Materials Management, Physical Distribution", Prentice Hall Inc, New Jersey, 1987.

[8] C. Arnold and W. Chiu, "IBM's Information Warehouse", DBMS 5(3), pp. 40-44, 87, 1992.