EDI - XML Standards and Technologies in the Agri-Food Industry

István Füzesi[1], *Miklós Herdon*[1]
[1] *University of Debrecen*

Centre of Agricultural Sciences Faculty of Agricultural Economics and Rural Development
Department of Business-and Agricultural Informatics

Abstract. Due to globalisation, the new technological developments and the complexity of food supply processes, the European food sector is increasingly becoming more complex. The consumers' trust in food, triggered and affected by a number of food crises, is low. Today, consumers increasingly expect safe and high quality food and demand information about the origin of their food. Also, the economic health of the food industry can be greatly affected by food crises; therefore, efficient and effective mechanisms are required to assist the food industry in tracking and tracing products along the food chain. In this paper, we discuss the criteria for an efficient and effective traceability system from an IT perspective (mainly data exchange) and we identify key requirements for ICT enabled traceability.
Keywords. Tracking and tracing, EDI, XML, Agro-Food Chain

Introduction

Increasing pressure by consumers, provoked by numerous food scandals, forced the European Union and national authorities to strengthen the regulations on food safety and traceability along food supply chains.

Commercial members of food chains are only partially prepared to implement such regulations and to fulfil the expected requirements. Major impediments are that existing IT-solutions are, in most cases, limited to single enterprises, there is an absence of authorities that control whole chains, small-scale enterprises at the supply and primary production level have to interact with large, (sometimes) multi-national groups of companies and food chains show significant fractures in data flows. A coordination and motivation problem crystallized as a key challenge in the provision of traceability and quality assurance. The coordination problem results from the division of labour in the value-added chain of agribusiness. This leads to the development of organizational intersections, which act as fractures in the information flow, with the result that the information flow is interrupted. The motivation problem is also a result of the division of labour. Diverse people and enterprises, with differing goals, are involved in the production, processing and distribution of food stuffs. The diverse actors must be motivated to ensure the complete and correct gathering and sharing of information needed to guarantee traceability, even if this does not contribute directly to their own interests (Theuvsen, 2003).

EDI (Electronic Data Interchange)

Electronic Data Interchange (EDI) is the process of using computers to exchange business documents between companies. Previously, fax machines or traditional mail was used to exchange documents. Mailing and faxing are still used in business, but EDI is a much quicker way to do the same thing. EDI is used by a huge number of businesses. Over 100,000 businesses have replaced the more traditional methods with EDI. This new system has a number of benefits; cost is one of them. Computer to computer exchange is much less expensive than traditional methods of document exchange.

The following components and tools are necessary for performing EDI:

- *Trade Agreement* - a legally binding trade agreement between a person and the trading partner.
- *Standard Document Format* - the standard agreed upon format for the document to be electronically transmitted.
- *EDI Translation Management Software* - software used to convert the document in the

application's format into the agreed upon standard format. For optimum performance the translation software should be on the same platform as the business application.

- *Communications Software* - a programming tool that enables the writing of communications protocols, or a separate application. It can be a module to the translator or a separate software application.
- *Modem* - a hardware device used to transmit electronic information between computer systems. The higher the baud rate, the faster the communications will be.
- *VAN* - stands for Value Added Network. A network to which one can connect to transmit data from one computer systems to another. One network can act as a gateway to another.
- *Point-to-Point* - a direct communication link from one computer to another. Some trading partners offer a direct connection to their EDI computer. Trading partners may opt for this method of communication instead of using a VAN.

EDI Standards

There are four major sets of EDI standards:

- The UN-recommended UN/EDIFACT is the only international standard and is predominant outside of North America.
- The US standard ANSI ASC X12 (X12) is predominant in North America.
- The TRADACOMS standard developed by the ANA (Article Numbering Association) is predominant in the UK retail industry.
- The ODETTE standard used within the European automotive industry

All of these standards first appeared in the early to mid 1980s. The standards prescribe the formats, character sets, and data elements used in the exchange of business documents and forms. The complete X12 Document List includes all major business documents, including purchase orders (called "ORDERS" in UN/EDIFACT and an "850" in X12) and invoices (called "INVOIC" in UN/EDIFACT and an "810" in X12).

The EDI standard says which pieces of information are mandatory for a particular document and which pieces are optional and give the rules for the structure of the document. Standards are generally updated each year.

EDI is a family of standards, including EDIFACT, HL7, X12, IATA, ODETTE, TRADACOMS and a plethora of others. They are all text file formats, which are structured so that each row of text begins with a marker that tells what kind of record that row contains, and each data item is separated by some unique character.

If a person wants to know which vocabulary a specific message type belongs to, generally the name or version number will give it away:

Clue	Vocabulary	Example
three digits	X12 transaction set	104, 810, 837, 850
six letters	EDIFACT message	DESADV, INVOIC, ORDERS
four digits	X12 release	3030, 4041, 5020
six digits	X12 release	003030, 004041, 005020
two digits then an 'A', 'B' or 'C'	EDIFACT version	93A, 99B, 04A

'D' or 'S'; two digits; and 'A', 'B' or 'C'	EDIFACT version	D93A, D01C, D05A
dot releases	HL7 version	2.3, 2.3.1, 2.5
file begins "UNA" or UNB"	EDIFACT document	UNA:+.? '
file begins "ISA"	X12 document	ISA:00: lots more :U:00304:000032123:0:P:*~
file begins "MSH"	HL7 document	MSH|^~\&||||||ADT|

Costs of EDI

Prices for EDI applications vary from free (for very simple one-function products) to several thousands of dollars for full-function applications. The final price paid depends on several things:

- *The Expected Volume of Electronic Documents.* Generally speaking, PC products cost less, but handle only a few documents and trading partners. Midrange EDI packages can be a little more expensive, but handle a much larger volume of EDI. If multiple documents or trading partners are anticipated, a midrange EDI system is a much better choice.
- *The Amplitude of the EDI Translation Software.* Some products look like a bargain, but as the EDI needs grow, hidden costs (such as having to purchase new transaction sets) suddenly appear. One may pay more for a program with an integrated mapper, but will avoid purchasing overlays and maps in the future.
- *Implementation Time.* Some applications are easier to learn and use than others. The more time spent in training, the more time it takes to get into production mode. If the time frame is tight, look for a translator that does not require training before implementation.
- Fees vary from software company to software company. Ignoring the hidden costs mentioned above, one can expect the following ongoing charges:
- *Maintenance Fees.* Most companies charge an annual maintenance fee that is usually a percentage of the translator's list price. This fee should include software updates, standards updates, technical support, and customer service.
- *VAN Charges.* The VANs bill for transmitting data similar to making a long distance phone call. Some also bill for connect time. A fast modem helps to lower transmission costs.
- *Mailbox Costs.* Most VANs charge a monthly fee for maintaining a mailbox on their network. Some base billing on the document (25 cents per document transmitted). Others charge based on the number of characters in each document.

Advantages of EDI

- **Save Money.** The cost of paper and paper processing is incredibly high compared to a properly implemented EDI program.
- **End Repetition.** If the trading partner wants a copy of a document, instead of calling, they simply check their mailbox. This results in a great time savings from not having to copy and fax/mail copies of business documents.
- **Save Time.** EDI also saves time over paper processing since the transfer of information from computer to computer is automatic.
- **Improve Customer Service.** The quick transfer of business documents and marked

decrease in errors allow one to do business faster and more efficient.

- **Expand Your Customer Base.** Thus with improved customer service, one can ultimately expand one's customer base. Many large manufacturers and retailers are ordering their suppliers to institute an EDI program. So, when evaluating a new product to carry or a new supplier to use, the ability to do EDI is a big plus.

Disadvantages of EDI

- **Too Many Standards.** There are too many standards bodies developing standard documents formats for EDI. For example your company may be following the X12 standard format, while the trading partner follows the EDIFACT standard format.
- **Changing Standards.** Each year, most standards bodies publish revisions to the standards. This poses a problem to EDI users. One may be using one version of the standard while one's trading partners are still using older versions.
- **EDI is Too Expensive.** Some companies are only doing business with others who use EDI. If a company wants to do business with these organizations, they have to implement an EDI program. This expense may be very costly for small companies.
- **Limit Your Trading Partners.** Some large companies tend to stop doing business with companies who do not comply with EDI. For example Wal Mart is only doing business with other companies that use EDI. The result of this is a limited group of people one can do business with.

Evolution of EDI

For more than 20 years, companies have been using Electronic Data Interchange (EDI) to transmit structured business documents like orders or invoices electronically. As opposed to paper-based communication, EDI is designed to make communication between different systems possible without media discontinuities. But although there undoubtedly is large savings potential, the use of EDI is by far not as widespread as one could expect. Forrester Research estimates that only about 5% of the companies which could profit from its use actually use EDI. The main reason is that especially small and medium-sized enterprises (SMEs) try to avoid the considerable setup- and operating costs of traditional EDI solutions. Therefore, the use of EDI is mainly reserved for large companies, and one of the main reasons for the introduction of EDI is usually pressure from larger business partners.

EDI systems are likely to be Internet-based in the future. While currently only very few enterprises use WebEDI, more than 50 percent of the enterprises plan to implement this technology in the future. A first step in reducing costs of EDI solutions is the use of the Internet with its existing communication infrastructure as a means of transportation for EDI messages. A number of different communication protocols can be used for the transfer of EDI messages over the Internet. Depending upon the task, the exchange can be made via FTP (File Transfer Protocol), HTTP (Hypertext Transfer Protocol) or SMTP (Simple Mail Transfer Protocol), while the data are encoded either with PGP (Pretty Good Privacy), S/MIME (Secure Multipurpose Internet Mail Extension) or SSL (Secure Socket Layer). Low cost in combination with the large number of emerging technologies for the Internet may now make a considerable contribution to the increase of EDI users and especially allow SMEs to participate in EDI networks. The obvious advantage of using the Web as a medium for EDI communication is that the only (client side) prerequisite is an Internet connection and a web browser. All communication uses the ubiquitous HTTP-protocol. Security issues can be addressed by using SSL, for example. Thus, all required infrastructure is most probably almost anywhere available without forcing the partners to invest hefty amounts of money. In this context, form-based EDI proves to be a good idea for large companies seeking ways of allowing their small customers to send their data in a standardized format.

The eXtensible Markup Language XML has the potential to be the data format of choice used together with the programming language of choice for the Web, Java, to enable the next

step in the evolution of EDI. The use of open standards can considerably reduce the time and money spent on implementing a solution. By avoiding proprietary formats, the danger of investment ruins is decreased and future-oriented solutions can be developed. Especially XML can contribute to the opening of EDI networks. While traditional EDI relationships are often long-term and highly integrated relationships, which are worthwhile only with are large number of transactions and for a long term, the willingness to invest into open, compatible IT-infrastructures is stronger at any point of the Value Chain, especially for smaller partners. Traditionally, the establishment of compatibility between different EDI solution-systems was achieved through deep integration of the EDI standard into the applications of the communication partners.

XML (eXtensible Markup Language)

Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere.

XML shares common origins with HTML and SGML. SGML or "Standard Generalized Markup Language" was issued as an international standard (ISO 8879) in 1986. This was intended for semantic markup that would assist computer cataloging and indexing. SGML provided flexibility that had not been available before and became very popular and was applied in many ways beyond the intentions of the original developers. It was, however, very complex and expensive.

About 1990, Tim Berners-Lee at CERN developed a new, simpler language that could be used in place of SGML. Thus was born HTML or "Hyper Text Markup Language." This was intended to be a simpler language that did not require expensive authoring tools. HTML succeeded beyond anyone's expectations but it lacked a certain flexibility that developers wanted. Various groups made changes and added extensions until HTML's roots had been mangled.

In the summer of 1996, a working group at W3C was formed to create a markup language that would combine the strength of SGML with the simplicity of HTML. The first official draft specification for XML was released in November 1996. XML version 1.0 became a W3C recommendation in 1998.The basic structure of XML is the document. This terminology, however, might cause one to think of XML as only a richer, more flexible HTML. It is richer and more flexible, but it can be so much more as well. Thinking of XML as a document allows you to see how it can be used for presentation of data. This presentation can be detailed and useful. Most browsers now handle XML for presentation. XML does, however, actually go beyond documents. It can be used for the communication of data as well. XML uses a flexible tagged structure that makes it more robust than a fixed record format for communication. Finally, XML can also be used to define the storage of data. The same flexible tagged structure can be used when storing data.XML standards are defined at breathtaking speed these days. It is also difficult to keep up with the various versions of those standards.

EDI - XML technologies in the agri-food industry

The ever increasing requirements for the level of tracking and tracing in turn require even more advanced IT techniques. Increasing the speed of a tracing operation is often the reason to store information about lots and tracing in an ERP system, but such a system can also be used to support RF scanning, to read barcodes and for EDI to communicate with clients and suppliers. Besides speed this also increases the efficiency of registration and this ensures that the data stored is correct and complete. Furthermore, an information system which has been set up well will also reduce dependency on people, because some of the knowledge is stored

in the information system itself.

There are three main aspects of the provision of information:

- Identification of the lot which is to be monitored. A lot has a unique identity created by the combination of item number and lot number.
- Registration and administration of the lot history: when was the raw material received from a supplier, when was this used in production, etc.
- Communication about the lot with other links in the chain. An important roll in this communication is played by the EAN standards for barcodes and EDI.

AGRO EDI Europe (AEE)

Since 1992, Agro EDI Europe is working on organization and standardization of electronic data interchanges on the agricultural and agro-industrial sectors. Today the association gathers about 250 members coming from several sectors (agricultural input, agro equipment, software editors, accounting centers, bank, insurance, logistics, storage, analysis laboratories).

Since 2001, within the Agro EDI Europe association, the economic partners of the farm, agreed to define a standard data-processing format of exchange for the feedback related to the data crop sheet: the DAPLOS message. When developing this "plot message", AEE works addressed the problems of follow-up of cultivation operations and computation of gross margin, and benchmarking of crop husbandry techniques. Later, AEE members turned to the domain of traceability, all tools enabling farm production management being marketed today as traceability solutions. AEE created the “plot message” (DAPLOS: Data Plot Sheet) which is a standard for describing information related to a specific cultural plot, in order to facilitate data exchange between various information systems. Software editors in France make many efforts to stick to this standard while developing programs and databases, mainly by implementing some export / import functions for writing or reading data according to AEE message. AEE DAPLOS message has focused attention on the relationships between the producer and service suppliers, and it does not especially address the Administration. If compared with AgroXML (see below) the data description is not included since it is an EDIFACT message. But the recent submission of the DAPLOS message to ISO authorities is an important step forward.

ADED / ADIS

LKV (Landeskontrollverband) Westfalen-Lippe designed a data dictionary for building up a unique German Internet database in the environment of ISO/SC19. As with the majority of such data dictionaries, it relies on items, entities, entity-item relationships and code sets. The data model has a multi-language capacity, is ready to work with XML and is available on the Internet. It could be integrated in existing users' systems.

AgroXML

Farmers are subject to a multitude of obligations concerning documentation of agricultural practices. agroXML is the result of a tight cooperation with producers of agricultural software and online service providers, which integrate agroXML into their software. agroXML introduces a standard, which facilitates data storage and exchange. agroXML is based on the international standard XML and consists of the agroXML-Schema and several contentlists. agroXML is available in an extended version 1.2. Several applications even in commercial software are existing since a couple of weeks.

AgroXML is a language that enables the description of agricultural data that will then enable a complete documentation of agricultural production processes. The objective of AgroXML is to allow information exchanges without redundancy between the different actors: land owners, farmers, advisory services, food industry. AgroXML will be available free of charge on the web, and will be platform independent.

An XML Schema defines electronic documents for data exchange. The agroXML schema is based on a model of the real-world processes in agricultural production. They are represented

in a tree-like hierarchy.

The current release of agroXML, Version 1.2, was published in March 2007 on www.agroxml.de and replaces Version 1.0 from May 2006. Schema development will be carried on in the English language in the future. Currently contained terms have already been translated but the translations have not yet been incorporated into Version 1.2. Other topics to be worked on will be further upgrading of geo-data functionality, as well as addition of elements for livestock farming and cultivation of vegetables and fruit.

Currently, different companies are implementing agroXML as a data exchange standard. Several examples are presented in the following text. Producers of agricultural equipment like e. g. Claas and John Deere support development of agroXML. Especially for the agricultural software industry, agroXML has a high priority.

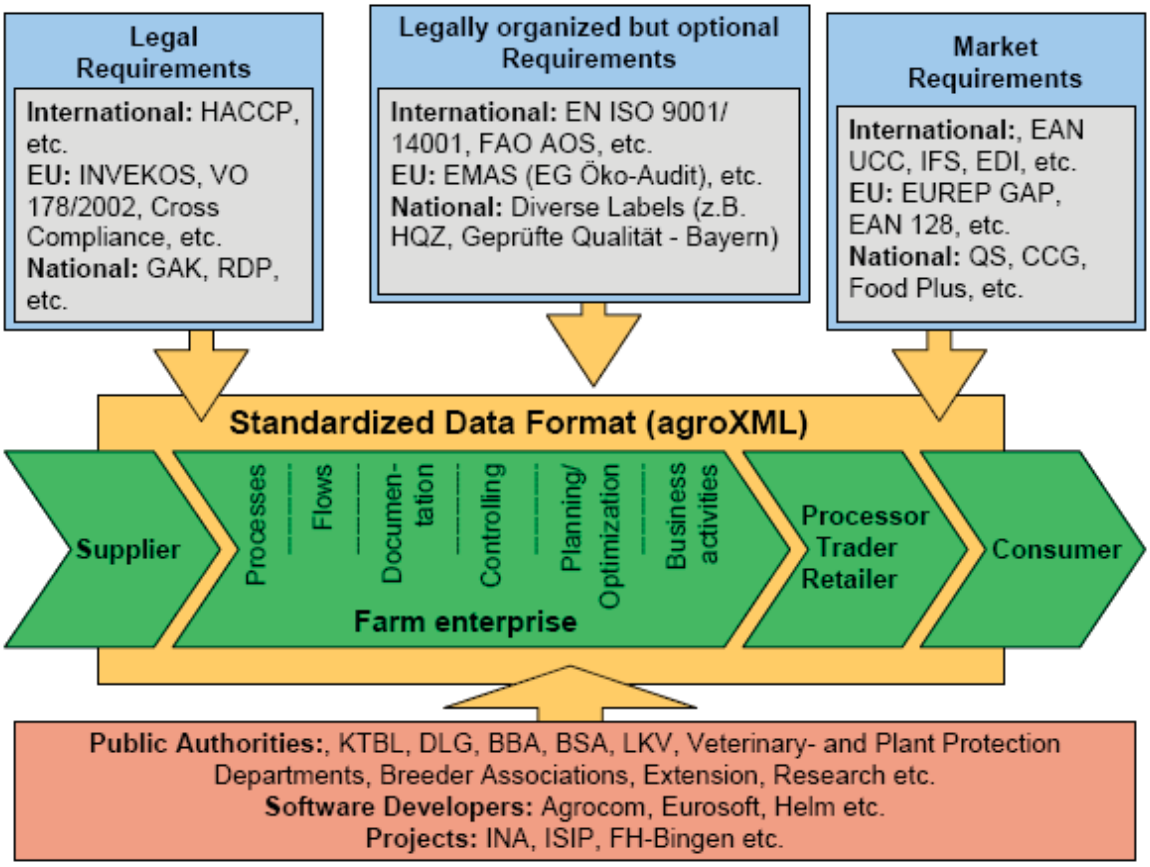

Figure 1. Data and information requirements in the Agribusiness sector (Source: According to Doluschitz, 2004)

In summary, the objectives of the development of agroXML are:

- the implementation of a generally acknowledged standardized data format to be widely used by all members of the agro-food chain,
- to avoid multiple data input at different levels of the agro-food chain and to minimize redundancy,
- to harmonize the discussion about the content and extension of the agroXML repository by involving experts from different origins on a neutral platform, and
- to increase and accelerate adoption.

The benefits of agroXML

Potential users of agroXML include anyone along the production and supply chain in the agricultural sector.

1. *Farmers*

In future farmers will not be tied to a specific collection or processing of data during the extensive conversion of documentation duties. There is no need any more to register data. agroXML facilitates an obstacle free communication with the agricultural administration, consulting services and software companies possible, without requiring an extra data input. Any already determined and documented data for production processes, which has been stored in any software, e.g. in a file for different types, are available at any time for further use.

2. *Consulting services*

Integrated plant production requires that agronomic measures have to be adapted to the habitat conditions of the single type as well as to the needs of the plant. For conversion there are many instruments, which are partly available at the place of business, or have to be purchased (soil testing, monitoring of stock, weather station, warning system, models for prognosis) The potential of the integrated plant production may be fully and most effectively used by the farmer, if consultants and extension specialists will in future provide access to individual and type oriented choice with the help of internet technologies. This way of acquiring consulting services makes precision farming particularly interesting (Doluschitz und Jungbluth, 2004a and b).

3. *Software companies*

The need of a standardized data exchange language like agroXML will increase and accelerate the development progress for agricultural software companies. Agricultural software is increasingly dependent on the data input from outside the place of business. On the one hand this concerns the updating and care of forms and demands for application of subsidies, and on the other hand the use of information concerning the means of the business. The corresponding updates require more staff and cause higher costs, which can be reduced substantially if this information will be available in a standardized data format via the internet. Increasingly the customers will also expect that the software with systems for the business und type oriented online consulting services can communicate without problems, which will be much easier when applying agroXML.

Examples of agroXML applications

1. *Field work calculator*

Currently in operation is access to the field work calculator of the KTBL (www.ktbl.de) from farm management systems of Helm and BASF. Using this tool, the user can calculate data concerning machinery costs and working time of flexibly compiled machinery combinations and import and automatically file them into his farm management information system. Usage of this system is also open to others.

2. *EurepGAP*

The example of the EurepGAP application shows how to transform agroXML documents automatically into instances, which conform to the requirements of an existing interface using XSL-Transformations. Involved in this use case is information, which is needed by EurepGAP for documentation of plant.

3. *KTBL-Tool geodata*

The application geo-data is developed by the KTBL in the course of a feasibility study supported financially by the BMELV and the Landwirtschaftliche Rentenbank. The prototypical application enables farmers to collect geospatial data from different providers in a simple manner and transform them into the agroXML format readable by farm management information systems.

4. Work commission using AGRO-NET of agrocom GmbH & Co. Agrarsystem KG

Today, the commission of a machinery cooperative happens using the telephone. Especially in seasons of high workload, when the commission is simply captured on a note placed on the desk, misunderstandings are possible. In the described use case, the electronic commission of services using agroXML has been developed. Using it, the farmer can enter his planned measure into his AGRO-NET management system and hand it on to the machinery cooperative at appropriate times.

Trace2p2

The P2P Project, co-founded by the European Community, aims to study, develop and test different methodologies and tools supporting the quick collection of information from different traceability systems, in order to define and develop a unified way to collect information automatically from the various traceability systems operating at the different companies. In this way each company can choose its own traceability system that best fits and on the other side customers, citizens and authorized persons will have a unified approach to inquire automatically for product details via Internet.

The main goals of the P2P project are to define a Methodology (Trace Methodology) and a Software Architecture that is the base to solve the key issues of integration. It means to conceive and develop a solution aimed at supporting quick collection of information from different traceability systems of companies belonging to a given agri-food value chain, giving reason to the whole value chain traceability. A pilot system will be developed for the swine value chain.

Technically the most important part of the software is a communication protocol to share the information, the TRACE-XML protocol. The *TRACE-XML* protocol defines the minimum set of information to trace each product batch.

The *TRACE-XML* protocol will represent the first step to define a standard to automatically access traceability information and documents of each company belonging to the agri-food sector. As for example, it could contain a subset of information and rules of the ebXML standard. The *TRACE-XML* protocol could represent an open standard that could address not only the swine, but also the whole agri-food sector.

For summarizing the main expected results from the project are:

- **Trace Methodology**: a methodology supporting traceability information collection in the companies (in particular how and what information);
- **Trace-XML protocol**: a minimum set of information to trace each batch and the related communication protocol;
- **Trace-SW**: a software tool supporting collection, archiving and management of information and documents related to the traceability;
- **Trace-Browser:** an ICT tool to browse the traceability information compliant with the TRACE-XML protocol.

Conclusion

Food quality has increasingly become an important issue in the last decade. That was primarily the result of a number of major food crises such as BSE and dioxin crisis that damaged seriously consumer's confidence on food industry. These striking examples of food deficiency had shown not only the vulnerability of food industry upon food quality but also revealed serious economic implications. It was clear cut that food crisis can not only affect consumer's health, but also food business's economic results and financial stability. The ever increasing requirements for the level of tracking and tracing in turn require even more advanced IT techniques. Increasing the speed of a tracing operation is often the reason to store information about lots and tracing in an ERP system, but such a system can also be used to support RF scanning, to read barcodes and for EDI to communicate with clients and suppliers.

Besides speed this also increases the efficiency of registration and this ensures that the data stored is correct and complete.

References

Bosak, J., 1997: XML, Java and the future of the Web. http://www.wiwi.uni-frankfurt.de/~tweitzel/paper/r-evolution/<FONT%20SIZE=2>http://sunsite.unc.edu/pub/sun-info/standards/xml/why/xmlapps.html%20

Brownell, D., 1998: XML and Java Technology, An Interview with Dave Brownell, http://www.webpaulo.com/html/xml_and_java.html

Curtis, C,,1996: EDI over the Internet: Let the games begin, in: Internetweek, Issue 627, September 9, 1996, http://www.techweb.com/se/directlink.cgi?CWK19960909S0076

Densmore, B.,1998: EDI vs. The new kids, in: Computerworld, April 6, 1998, http://www.computerworld.com/home/Emmerce.nsf/All/980406edi

Doluschitz, R., 2004. Der Beitrag der Informationstechnologie zu Produktionsmanagement, Qualitätssicherung und Rückverfolgbarkeit in der Agro-Food-Chain - Notwendigkeit, Strategien und Perspektiven: In: FAT-Schriftenreihe, Agroscope FAT Tänikon/CH, 2004, H. 59, 5-25.

Doluschitz, R.; Kunisch, M.; Jungbluth, T.; Eider, C., 2005: agroXML - A Standardized Data Format for Information Flow in

Agriculture, In: EFITA/ WCCA 2005, Portugal, Book of Abstracts, Juli 2005, S. 26

Kunisch, M., Frisch, J., Martini, D., Böttinger, S., 2007: agroXML – a standardized language for data exchange in agriculture In: EFITA/ WCCA 2007, Scotland, Book of Abstracts

Huggins-Cha, S., 2007: EDI http://ksi.cpsc.ucalgary.ca/courses/547-95/seanh/edi.html

Marcella, A./Chan, S.,1993: EDI Security, Control, and Audit, Norwood 1993

Segev, A./Porra, J./Roldan, M.,1997,: Internet-Based EDI Strategy, working paper 97-WP-1021, http://haas.berkeley.edu/~citm/wp-1021.pdf

Theuvsen, L. 2003: Rückverfolgbarkeit von Lebensmitteln: Herausforderungen und Lösungsansätze aus organisatorischer Sicht. In: Berichte über Landwirtschaft, Heft 4/2003, S. 555 – 581.Waltner, C.,1997,: EDI Travels The Web - EDI over the Web offers companies cheaper E-commerce and messaging via browsers, in: Internetweek, Issue 668, June 16, 1997, http://www.techweb.com/se/directlink.cgi?CWK19970616S0066

Westarp, F. v./Weitzel, T./Buxmann, P./König, W., 1999: The Status Quo and the Future of EDI, forthcoming in: Proceedings of the 1999 European Conference on Information Systems (ECIS'99), http://caladan.wiwi.uni-frankfurt.de/IWI/projectb3/deu/publikat/edi/index.htm