Coordination and Exchange of XML Catalog Data in B2B

Joerg Leukel, Volker Schmitz, Frank-Dieter Dorloff

University of Essen
Department of Information Systems
Essen, Germany
{joerg.leukel|volker.schmitz|dorloff}@uni-essen.de

Abstract

The transmission of electronic product catalogs using e-business standards belongs to the first and furthest common applications of XML in B2B e-commerce. Despite the application of standardized catalog formats (e.g. cXML, xCBL) often a relevant amount of coordination and communication between the catalog-producing and catalog-importing companies is still necessary. If single standards offer degrees of freedom or their specifications contain interpretation clearance, errors in catalog exchange can occur. This aspect is of a special importance, since thereby someone has to intervene in the exchange process manually; this contradicts the automation paradigm of e-business. With consideration of the import errors, catalog creation and catalog import must be repeated, until the catalog is regarded as completely valid by the target system. In consequence the exchange processes are little automated and costly as well as time-intensive. This paper discusses the design of catalog data exchange and suggests a three-stage improvement concept. Its basis is a close look at current practice in catalog exchange and data characteristics that determine the exchange processes. Our approach addresses the field of e-business standardization, since commercial XML standards have already started forming a de-facto data infrastructure for e-business in general. Therefore we will analyze several catalog standards and ask what mechanisms they provide to design and implement efficient exchange processes.

Keywords B2B, E-Business, E-Catalogs, E-Procurement, Standardization

1. INTRODUCTION

Given the rapid growth and success of web-based procurement systems, e-catalogs gain an outstanding importance. Product catalogs form the basis for buying decisions and the release of order transactions. Thus they are a prerequisite for electronic markets [1]. In contrast to B2C, catalog usage in B2B is characterized by the fact that data of the catalog-creating enterprise is imported into an information system (target system) of the catalog-receiving enterprise. Meanwhile XML-based catalog formats became generally accepted [2]. Despite the application of standardized catalog formats (e.g. cXML, xCBL) often a relevant amount of coordination between supplier and buyer is necessary. Especially in the initialization phase, when the first exchange between two partners is established, a lot of adjustments regarding syntax, contents and quality of the transmitted data have to be made.

Despite the application of standardized catalog formats (e.g. cXML, xCBL) often a relevant amount of coordination and communication between the catalog-producing and catalog-importing companies is necessary. Especially in the initialization phase, when the first exchange between the partners is established, there are a lot of adjustments regarding syntax, contents and quality of the transmitted data to be made.

A starting point for the improvement of exchange processes is extending the XML catalog standards in such a way that they support the coordination and the exchange more widely by providing an appropriate process model and additional business messages. This approach is pursued by our paper, which is structured as follows: In Section 2 we will describe, how catalog exchange processes are carried out today and which demands the market partners make on these. On this
basis we will identify in Section 3 several characteristics of catalog data that must be considered during the process organization. With knowledge of the operating conditions a three-stage improvement concept is developed in Section 4; it can be used for the extension of XML catalog standards. Subsequently, we will check empirically how the components of our concept are already implemented in industrial catalog standards (Section 5). A summary of the results will close our paper.

Research literature shows two main working areas regarding e-catalog data. The first area deals with the integration of different catalog data formats. A major question is how to solve the mapping problems. Several techniques can be applied: low-level XSLT transformations [3], RDF data models, e.g. [4], and enhanced Naive Bayes classifications [5].

Approaches for a semantic integration of product classification systems form the second area. Classification systems are seen as product ontologies that enable a common and accepted communication in the respective product domain [6]. The specification of ontologies takes place using concepts of knowledge representation. An integration approach is presented in [7]: the system imports classification system data using a Wrapper. Supply of this data in a standardized format is not demanded. In [8], a similar, integration-oriented path is taken. [9] proposes an information retrieval approach.

While the goal of most related work is the integration of existing catalog data formats, our paper addresses the standardization and improvement of catalog data and catalog documents. We argue that industrial XML catalog standards are not sufficient to meet all requirements of catalog data exchange.

2. XML CATALOG DATA in B2B

Contrary to most data stored in operational information systems catalog data is exchanged between enterprises by definition. Differently than data e.g. in controlling or sales, catalog data does not remain within the enterprise boundaries, but is supplied to customers and is used by them. A reason for this is that procurement in B2B is made ever more strongly through buy-side systems and e-markets. Buy-side systems are e-procurement systems being operated by large buying enterprises. Their aim is to optimize inter- and intra-organizational purchase processes. E-Markets bring several suppliers and customers together. In addition, catalog data exchange is not limited to the relationship supplier-buyer. In many branches of industry catalog data is exchanged along the entire supply-chain, e.g. manufacturer – wholesale – industry. On the other hand sell-side systems, typical e-shops that represent the products of only one supplier, lose their former importance [10]. Here we look at catalog data exchange as the transfer of catalog data into target systems; however we do not cover the execution of queries on distributed catalog data so that ad-hoc virtual catalogs are built [11].

Catalog data possesses a substantial meaning for suppliers. It describes their products and is an instrument for differentiation between the competitors. To that extent high-quality catalog data can be regarded as a valuable economical good, which contains bundled know-how of products. The task of catalog creation often makes the introduction of new or the extension of existing information systems necessary. A reason is that catalog data is stored in different and distributed operational information systems. Likewise the relevant data is administrated by different organizational units. Established concepts and enterprise-wide data models for the catalog data management are often missing. In addition catalog data management has to integrate several sources and must be able to create catalogs in any XML standard fast and economically.

The target systems must be able to import any XML catalog document. Especially for market places, which process hundreds of supplier catalogs, the catalog import is a key task; particularly since it cannot be assumed that all catalogs use the same format and that their quality is evenly high. Therefore catalog data passes through a staging process that covers different technical and semantic checks, operations and release steps. At the end stands a checked catalog, which is ready for the use in operational systems. Defined staging processes are useful, because a syntactical correctly created catalog does not guarantee that the catalog corresponds also to content requirements. Concerning this, a statement can be made only in dependence on the respective, often customized requirements.

If single standards offer degrees of freedom or their specifications contain interpretation clearance, errors in catalog exchange can occur. This aspect is of a special importance, since thereby someone has to intervene in the exchange process manually; this contradicts the automation paradigm of e-business. With consideration of the import errors,
catalog creation and catalog import must be repeated, until the catalog is regarded as completely valid by the target system. In consequence the exchange processes are little automated and costly as well as time-intensive.

3. CHARACTERISTICS OF CATALOG DATA

Catalog data indicates a set of characteristics, which lead altogether to high requirements on catalog data management and its supporting information systems. These characteristics influence substantially the inter-organizational catalog data exchange.

**Change Frequency:** Catalog data has mostly a master data character (e.g. technical product features), however specific data changes over time (e.g. prices) or can be determined only time-dependently. Therefore modifications of catalog data have to be communicated to the catalog users, or certain data has to be queried from the supplier if necessary. A simple catalog transfer en-bloc is not sufficient, but also catalog updates and requests for updates should be implemented [12].

**Multi-Dependency:** We must consider that catalog data does not mean the data of only one specific catalog. Rather catalog data in its whole represents the quantity of data from which catalogs can be created. The creation of specific catalogs refers to an important characteristic: Each catalog possesses a validity, which can be determined by a set of parameters. Beside the customer, this is the validity period, the currency of prices and the language of language-dependent data. In accordance with this multi-dependency concept each catalog can be described and identified by a tuple of parameters [13]. It has to be considered that also multi-supplier catalogs, catalog documents with product data of several suppliers, as well as multi-vendor catalogs that contain customized data of several buyers, should be transferred. The catalog exchange must guarantee that the right catalog is transmitted in the required format to the addressed recipient.

**Buyer-specific Data:** The most important dependency parameter for catalog data is the buyer. This dependency refers both to syntax and semantics of the data [14]. Usually the buyer sets the catalog format and required content. Buyer-specific modifications or extensions of the standard formats have to be considered, too. Regarding data contents the requirements can be divided into such, which relate to the transfer of customized data (e.g. prices, article descriptions and numbers), and those, which declare in accordance with the standard optional contents as mandatory contents (e.g. product figures, danger property classification). To sum it up, even when suppliers and buyers use a standardized catalog format a bilateral agreement – often determined by the buyer – is mostly necessary.

**Content Complexity:** Part of the catalog data is not only the data that describes products directly (e.g. identification, names, features); rather catalogs build a system of high complexity. It consists of meta data, product data, price data, classification data and structure data, which are represented by more or less complex XML data models. Compared with other business documents the complexity of catalog data is higher. This shows up clearly in the respective specifications of the standards, which partly cover over 100 different data elements, and its transformation into a relational model requires over 20 entity types (numbers for BMEcat standard). The complexity during the catalog creation process must be controlled by the supplier. In particular a broad spectrum of different catalogs for the suppliers’ customers can accompany the content complexity. On side of the catalog-processing system a semantic validation must consider many aspects in order to give a reliable statement about the catalog quality.

**Standardized Formats:** The exchange of catalog data makes use of standardized XML messages. It is not expected yet that a single, generally accepted format will prevail worldwide and for all branches of industry. For this reason suppliers and market places must be able to support the most important formats at the same time [15]. The standards define on the basis of a vocabulary for catalog data several business messages, e.g. Catalog or CatalogUpdate. From the view of the standards a catalog exchange process consists of a business message or a sequence of relating business messages. The organization of the catalog exchange processes depends substantially on the catalog standards and the business messages supplied by them [16].

**Document Size:** Making more difficulties is the size of the data that must be transmitted and processed. Extensive catalogs with up to hundred thousand products and attached multimedia objects can be, not least because of the XML tags, several hundred MB large [17]. This leads to the fact that asynchronous communication mediums (E-Mail) or offline mediums (CD-ROM) are used. Processing and validating a document can take several hours, even with the...
application of optimization techniques. Especially in case of the first catalog transfer the syntactical and semantic validation is time intensive. Thus process times are high and they can multiply if processing errors occur.

4. AN IMPROVEMENT CONCEPT FOR CATALOG DATA EXCHANGE

Altogether we see that the empirical process characteristics, introduced in Section 2, can be attributed to typical characteristics of the transferred data. The described situation in catalog data exchange shows several lacks and inefficiencies, which concern both the catalog suppliers and the catalog-processing enterprises. If we want to answer the question, by which measures the processes can be improved, we must consider organizational as well as technological aspects. In the following we develop an improvement approach that is based on two principles. On the one hand we set the priority goal of meaningful automation. On the other hand we understand the area of e-business standards as the primary field of action.

The paradigm of meaningful automation leads to system-to-system communication. Applied to e-catalogs this means that the involved information systems that create and import catalog data should communicate together. Most catalog systems used today are still far away from that. The communication is hardly bilateral, messages are only transferred one-way to the target systems and no situation-dependent feedback is given by these. We concentrate on the area of e-business standards to reach a significant improvement and automation of exchange processes. The reason is that available commercial XML standards (and the business data standardized by these) form a de-facto data infrastructure for e-business in general. Here is a distinction between document-oriented standardization of business messages and the standardization of frameworks for the transport of messages useful [18].

In the further we ask, what contribution can make the instrument of catalog data standardization in order to improve the catalog exchange processes. For answering this question a look at the characteristic differences between B2B and B2C relationships is helpful. They are pointed out in [19] on the basis of three criteria:

• Interaction between information systems is essential.
• The business content is diverse and complex.
• The control mechanism ranges from one-sided to peer-to-peer relationships.

Now we will use these three aspects for the development of our concept by defining three appropriate development areas with knowledge of the specific characteristics described in Section 3.

The first area covers the creation of a process model, which can be used by catalog systems for their communication to a large extent. The main requirement is that it is laid out as a request-response model and so that it contains not only one-way messages.

The second area addresses the content complexity by introducing extended possibilities for the specification of catalog contents. By catalog specification we mean the explicit specification of the contents provided or requested by suppliers and customers; it adds specific meta data to the document headers. The catalog specification should point out between the involved partners, which data are provided or requested and how the catalog standard is concretized and limited if necessary.

The third area applies the concept of jointly design of exchange processes to the area of catalog validation, and introduces a message type that informs the catalog creator about the import status and errors.

The process model must cover all catalog-relevant communication relationships between the enterprises involved, and defines message types and valid sequences of messages (processes). If all data flows are modeled, then the model is complete. Thereby a main requirement is independence from business models and frameworks. Business model independence says that both direct relationships (supplier-buyer) and intermediate relationships are subject of the process model. Framework independence means that those aspects of communication, which are not specific for catalog data exchanges, are not modeled, since they belong to frameworks; e.g. security, message handling and transportation [20].
In business transactions catalog data is subject of the information phase. Whereas the sequence of exchanged messages can be long in the agreement and execution phases (e.g. quotation, order, order confirmation, delivery note, invoice, payment), the message sequences in the catalog area are shorter, i.e. the dependencies between individual business messages are smaller. Frequently the communication is of a request-response kind. This can be shown at three catalog data processes, which occur at different points in time in the information phase. One process refers to the first transfer of the catalog, where the whole of the necessary data has to be transferred. If necessary, an explicit catalog request by the buyer (or e-market) can precede this transfer (Request for Catalog).

The next process deals with the updating of a catalog already transferred (Catalog Update). Under the term update we summarize changing, deleting and adding data. We must differentiate whether the catalog creator starts the update or the catalog user requests an update (Request for Catalog Update).

A third process is started, if a catalog user requests additional data that is not part of the catalog (e.g. customer prices on an open market place) or cannot be part of it (e.g. delivery time, real-time prices). This process is initiated by a Request for Quotation. The response (Quotation) can be different. On the one hand, the response can be generated automatically or the supplier has to edit the response manually. On the other hand, the response can be a rejection, acceptance or even modification of the request.

The process model is presented in figure 1 and contains an additional message type CatalogImportResponse, which will be introduced later.

A process model is a prerequisite for each catalog data exchange. However, it is yet not expressed which data should be transferred in an individual business relationship. This is the subject of a catalog specification that describes the catalog content explicitly. Catalog standards offer options and contain interpretation clearances. In order to enable a smooth import of catalog data this unclearness should be clarified before the data exchange. Need for coordination results mainly from different business process rules and the required data; though it arises from technical possibilities and basic conditions of the catalog systems involved as well.

The catalog specification is implemented by a set of meaningful meta data that is added to the document header. It should be used constantly in the messages RfC, RfCU, Catalog and Catalog Update:

- Validity: Hereby the supplier determines the validity of the catalog documents by naming supplier, buyer, validity period, currency and language in accordance with the multi-dependency concept. Likewise the recipient can express requirements in a request message.
• Process Model: If a supplier transfers a catalog in a standard format, then it cannot be assumed that its catalog system is able to process and produce all message types offered by the standard. The same applies to the catalog-requesting enterprise. The supported message types are included in the meta information.

• Content Specification: The catalog recipient requests certain contents in the context of a catalog standard, which are necessary due to company- and application-specific requirements. For example, Intranet-based procurement systems demand product figures, description texts without abbreviations and customer-specific prices. Further requests are the classification and description of products with specific features, the definition of semantic relations between products, and alternative keywords for the product search. Content specification can be extended to all data covered by the standard. Beyond that, requests can limit the range of the products by requesting those that belong to certain product groups (e.g. office supplies).

• Additional Data: If the data element definitions of the catalog standard are not sufficient, they can be extended in an individual agreement. In particular thereby customized information can be transferred, e.g. ERP data like cost centers, buyer groups.

• Import Restrictions: Turned around, it can occur that the catalog system used by the catalog recipient is not or only partially able to process certain data, though this data is defined in the standard format. Or some data is not needed or welcome at all. These restrictions of the catalog-systems concern frequently field lengths, which are limited e.g. article number and names.

The described requirements lead to new or modified specifications of XML data elements. Independent of content wise criteria and reasons, these specifications can be attributed to the following operations:

• Changing optional to mandatory data elements

• Excluding data elements

• Limiting the domains of data elements

• Limiting the domains of attributes

• Adding data elements

In view of the vast possibilities of specifying a catalog individually but in accordance with a standard, it must be guaranteed that the standard is not violated, e.g. domains of the standard specification should not be hurt and data elements should not be interpreted in a different meaning.

The two concepts process model and catalog specification cooperate as the different messages, which form the process, can make use of the instrument of catalog specification. So the request messages can determine exactly, which data has to be returned. This shows up well in the message type Request for Catalog that for example can contain the following specifications: “The default currency should be EURO and the default language should be English. The process model covers RFIC, Catalog, RFICU and Catalog Update. All product prices should be customer net prices with validity starting from 2002-01-01. All products have to be classified in accordance with the classification system eCl@ss 4.0. The length of article numbers is limited to 16 characters.”

The supplier should repeat this specification in the message Catalog and if necessary add further aspects. Thus the catalog recipient can control through a comparison of the request specification with the received specification whether the supplier accepted all requirements, and is in case of differences able to react accordingly.

Finally, the error handling during the catalog validation remains as the third area. This area, which is covered by the message type CatalogImportResponse, has to be put out separately, since the interaction of the catalog systems involved is very high. The aim is to inform the catalog creator about the results of the catalog import in a qualified manner, and – in case of an error – about the reasons of failure. In an ideal case, the catalog-creating system imports the response message and is in a position to recover certain error types automatically. Other error types will make manual interventions by a catalog system operator necessary.

An import report should contain the following information: Which catalog was imported (in accordance with the multi-dependency concept) ? How was the catalog imported (enterprise, software system, contact person) ? When was the
catalog imported (beginning and end time) ? Total status: Ok, Import with errors, Import aborted; List of occurred messages.

The list of import messages contains all relevant events, which occurred during the import. According to their type the messages can be split into different categories, which reach from notes over warnings up to serious errors, which led to the abort of the import. For example the absence of requested graphic data may be a minor problem, whereas the violation of the syntax of the standard (e.g. false data type, missing mandatory items) must be handled in a special way. The message list represents the result of a syntactic and semantic validation that can be executed partly by XML parsers under application of the formal XML specification of the respective standard. A meaningful report should contain an error statistics and for each error the type and line in the document as well as a note or explanation of the cause.

5. ANALYSIS OF XML CATALOG STANDARDS

On the basis of the improvement concept introduced above we can now examine selected catalog standards, ask what parts they already implement and determine, which areas have the smallest support so far. The selection is limited to the most important, horizontal standards. The selection covers the following standards:

- BMEcat is a genuine catalog standard. It is completed by the transaction standard openTRANS that contains among other messages the RfQ and Quotation messages [21].
- cXML is the standard data exchange format used by the e-procurement solutions of Ariba. The focus is here on the supply of formats for catalog-based order processes [22].
- OAGIS contains over 200 XML transactions for business documents. It will be integrated into the ebXML framework and cover the document level [23].
- xCBL is an extensive collection of XML business documents developed by CommerceOne [24].

The first area of analysis concerns the implemented process model, thus the seven message types (table 1).

<table>
<thead>
<tr>
<th></th>
<th>BMEcat 1.2</th>
<th>cXML 1.2</th>
<th>OAGIS 7.2.1</th>
<th>XCBL 3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Request for Catalog</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>2. Catalog</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3. Request for Catalog Update</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>4. Catalog Update</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>5. Request for Quotation</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>6. Quotation</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>7. Catalog Import Response</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 1: Process Models in XML Catalog Standards

OAGIS covers with six message types most of the requirements, but none of the standards implements the processing concept completely. Additionally it is noticeable that the three other standards do not specify Request for Catalog or Request for Catalog Update messages. Thus catalog communication remains limited to one-way transfers. cXML is the only standard that contains a mechanism to report the import status to the catalog creator. To accomplish this, cXML transmits a CatalogUploadRequest first, to which the catalog document is added as an attachment. Thereupon it receives feedback messages as long as the final status is not reached. The range of possible status reaches from Accepted over Published up to HasErrors. A HasErrors status signalizes that the document is syntactically not correct, i.e. it does not correspond to the cXML specification. However a semantic check does not take place and it is also not possible to check additional requirements.
Within the second area we examine, which possibilities exist to use meta information for an extended catalog specification. The results are presented in table 2.

In the case of the multi-dependency concept no standard provides the possibility to set the parameters (supplier to language) individually. BMEcat and xCBL realize this concept furthest, cXML not at all. The classification system, according to which the products are classified or should be classified (within Request), cannot be specified in any of the examined standards. Instruments for concretizing the specification concerning data elements, attributes and scopes are only supplied by OAGIS, though it is limited to the modification of optional to mandatory fields. To achieve this a GET_ECATALOG message is transmitted, which includes exactly the data elements that have to be transferred in any case.

<table>
<thead>
<tr>
<th></th>
<th>BMEcat 1.2</th>
<th>cXML 1.2</th>
<th>OAGIS 7.2.1</th>
<th>xCBL 3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Supplier</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Buyer</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Area</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Time Interval</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Currency</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Language</strong></td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Classification System</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Extended Definitions</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Mandatory Fields</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Ignore Fields</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Limit Attributes</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Limit Domains</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2: Catalog Specification Instruments in XML Catalog Standards

6. CONCLUSION

In this paper we have discussed the design of catalog data exchange and suggested a three-stage improvement concept. The basis of our concept was a close look at current practice in catalog exchange and data characteristics that determine the exchange processes.

Our approach addresses the field of e-business standardization, since commercial XML standards have already started forming a de-facto data infrastructure for e-business in general. Therefore we analyzed several catalog standards and asked what mechanisms they provide to design and implement efficient exchange processes. We found out that none of the four selected industrial standards realizes all design parameters of our concept. All things considered, we must say that the standards do not meet the requirements, especially if we look at requests of e-markets and content hubs. This underlines our conviction that further research and standardization must be done to come to universal and accepted business documents and processes. Our work is complemented by the development of reference models for specific catalog data, for instance price models [13] and classification systems [25].

References


