

FP 53 Integrated Specification and Documentation based on XML Schema Annotations

Volker Schmitz, Joerg Leukel, University of Duisburg-Essen, Institute for Computer Science and Business Information Systems, Essen

Numerous XML-based formats have been developed to realize intra- and inter-organizational data exchange. This is especially true for e-business standards that support business processes. Initially started as small XML messages covering only very specific transactions, today often extensive sets of message types are developed. Extending the usage of XML technologies to further, more sophisticated and critical applications leads, however, to a higher complexity of the underlying XML formats. As a result, standards adopters face large numbers of message types, data elements and data types. Providing an unambiguous, consistent and user-friendly documentation in an efficient way is the major challenge in e-business standardization. One of the main tasks is to ensure that the formal specification specified in XML Schema (XSDL) and the textual documentation are consistent.

Next we present a concept for enriching XML schemas with annotations to combine the formal specification with the textual description. This approach allows the generation of the documentation fully automated and without inconsistencies. Looking at major standardization initiatives, specification and documentation are often developed separately from each other. One reason is the lack of powerful software tools for engineering XML message types and producing customized documentations.

Information Requirements

The central part of the documentation of is the description of the vocabulary. Its items are described, for instance, by its data type, sub elements, and attributes. Figure 1 shows respective meta data for XML elements and attributes.

Meta data	Content
Name (formal)	Name of the item as in the XML schema (e.g., PRODUCTID)
Name (textual)	Written out name of the item (e.g., Product Identifier)
Short description	Short description
Long description	Detailed description; may include illustrations or tables
– Examples	Examples that explain the usage (XML code)
– Changes per version	Textual description of the changes due to new versions
Graphical representation	Graphical representation of the item and its sub items (e.g., elements with sub elements and attributes)
Data type	Data type of the item
Data type facet	Facet of the data type, i.e. field length or minimum/maximum values
– Values (permitted)	Domain restriction with permitted values (enumeration)
– Values (predefined)	Domain restriction with predefined values (pattern)
Default value	Default value of the item
Language dependency	Indication whether the content of the item depends on the used language and therefore can be specified multiple
Usage (super ordinate elements)	List of elements where the item is used
Sub elements	List of sub elements of the item
Version number	Number of the version in which the item was changed last
XSD / DTD extract	Extract of the XML schema with the definition of the item
Cardinalities	Cardinalities of the sub elements
Attributes	List of attributes of the element

Fig. 1: Meta data for XML elements and attributes

Information Sources

For generating the documentation based on the specification automatically, all relevant information has to be represented machine-readable. We categorize this information as follows (see Figure 2):

- The darkly marked information is already contained in the XML schema of the standard. A prerequisite is, however, that the respective modeling concepts of XSDL are actually used.
- The medium marked information can be derived from the XML schema.
- The brightly marked information can not be extracted from the XML schema; hence this information has to be supplied additionally.
- The non-marked information is not relevant for the respective item type (e.g., attributes do not have attributes).

Meta data	Element	Attribute	Data type	Value
Name (formal)				
Name (textual)				
Short description *				
Long description				
- Examples				
- Changes per version				
Graphical representation				
Data type				
Data type facets				
- Values (permitted)				
- Values (predefined)				
Default value				
Language dependency				
Usage (super ordinate elements)				
Version number				
XSD / DTD extract				
Sub elements (see element)				
Cardinalities				
Attributes				

Fig. 2: Mapping of meta data to item types (4) and sources of information (3)

Information Distribution

From the requirements of reusability and distributed development follows the need for storing the information specified above as separately as possible in independent files. The distribution of the information takes place dependent on the scope. We distinguish the following groups of information (see figure 3):

- Element-specific information serves for the description of the elements (and data types) being part of the vocabulary.
- Context-independent element-specific information describes characteristics that are independent from the use of the element (context, scope); for instance, the element name must not change at all.
- Context-dependent element-specific information describes characteristics that are only valid in its context, thus they may be different in other contexts. The context is determined by the super ordinate element to which the element is assigned.
- Element-spanning information refers to the standard generally or to components of it (e.g., introduction, legal notes etc.).

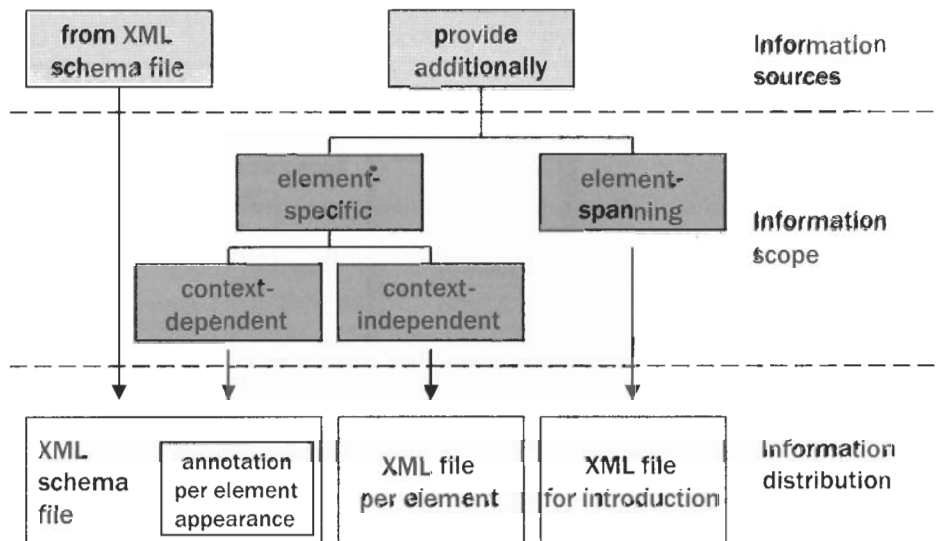


Fig. 3: Sources of information, ranges of validity and distribution

Schema Annotations

Figure 3 has already shown that the element-spanning information as well as the context-independent information is stored in a single XML document per each element. All context-dependent information is bound to the structure specified in the XML schema. Therefore, it is not meaningful to build a second, parallel structure. Thus, context-dependent information can be inserted directly into the XML schema. Such an XML schema, which has been enriched by semantic information, is called “annotated schema”.

There are two alternatives for enriching XML schemas: On one hand, the already defined elements can be extended by additional XML attributes of a new name space. On the other hand, the “xsd:annotation” can be used which is explicitly intended for providing documentation and application information within XML schemas. We choose the latter because it draws a clear dividing line to the ordinary “xsd:element”s.

Evaluation

Our concept has been evaluated by implementing a respective software tool and using this tool for the standardization project BMEcat 2005, which is the newest version of the BMEcat standard for the exchange of electronic product catalogs in B2B e-commerce (www.bmecat.org). The BMEcat vocabulary consists of 401 elements, the documentation totals 697 pages.

Figure 5 describes the generation process and shows the involved file types and used software tools.

The following improvements were achieved during the standardization process (compared to the previous BMEcat version):

- The time, which had to be invested into quality assurance, could be significantly reduced, since inconsistencies between the formal specification and the non-formal documentation were prevented in advance.
- The developers could concentrate on the technical specification and the content wise description, since manual, page-oriented editing and

