EnviroInfo 2002 (Wien)
Environmental Communication in the Information Society - Proceedings of the 16th Conference
 ISBN: 3-9500036-7-3

Standardisation of XML-based DTDs for Corporate Environmental Reporting: Towards an EML

Christian Lenz, Ralf Isenmann[1], Jorge Marx-Gómez, Mario Krüger[2], Hans-Knud Arndt[3]

Abstract

Discussions concerning companies' environmental aspects gained considerable importance and challenged business in recent years. Comprehensible and authentic as well as customised corporate environmental reporting requires substantial quantities of relevant information. In order to automate corporate environmental reporting, the presented approach focuses on structuring and standardising environmental documents using the eXtensible Markup Language (XML). In a more detailed fashion, the contribution is dealing with the harmonisation of three current XML-based document type definitions (DTDs) proposed for internet-based corporate environmental reporting. From an academics' point of view, this harmonisation is thought to be a considerable effort: firstly, for contributing to the discussion of standardising corporate environmental reporting taken as a whole, and secondly, for stimulating the processes towards shaping an Environmental Markup Language (EML). From a practitioners' perspective, such a standardised XML-based DTD contributes to employ Internet technologies and Internet services productively for corporate environmental reporting. Reporting companies and their target groups are enabled to exploit the huge opportunities and media-specific technical benefits efficiently. The paper is structured in four parts: The first part deals with the design process used to develop each DTD. In the second part, the three DTDs are presented in a more detailed fashion. Based on this, the third part refers to the process of harmonisation of the proposed DTDs. The paper ends with a conclusion.

[1] University Kaiserslautern, Department of Business Information Systems and Operations Research (BiOR), Gottlieb-Daimler-Str., P.O. Box 3049, 67653 Kaiserslautern, Germany, email: {lenz|isenmann}@bior.de, Internet: http://www.bior.de

[2] Otto-von-Guericke-Universität Magdeburg, Faculty of Computer Science, Institute for Technical and Business Information Systems, P.O. Box 4120, 39016 Magdeburg, Germany, email: gomez@iti.cs.uni-magdeburg.de, Internet: http://www-wi.cs.uni-magdeburg.de

[3] Humboldt University of Berlin, School of Business and Economics, Institute of Information Systems, Spandauer Str.1, 10178 Berlin, Germany, email: arndt@wiwi.hu-berlin.de, Internet: http://www.wiwi.hu-berlin.de/iwi/

1 Design Process

All three proposed DTDs were prepared autonomously and published more or less simultaneously. Due to the independent preparation from each other, at first glance, one may expect that analysing the DTDs in a more detailed fashion would show a rather different result. However, on top level all three DTDs appear almost similar, and even in more detail, the first two are looking quite similar to one another. Moreover, the really strong similarity has its basis in the same methodology used, on which the DTDs rest. This methodology represents a sophisticated and comprehensive procedure for designing DTDs divided in four stages (Schraml 1997, 122):

- Definition of the main target (stage 1): What are the primary purposes the documents mainly serve and what kind of requirements have to be taken into account.
- Identification of possible semantic components (stage 2): According to the defined target, several resources have to be analysed for identifying the pool of all possible semantic components (PSCs) the DTD should contain.
- Selection of relevant semantic components (stage 3): Of the pool of PSCs, a catalogue of actually relevant semantic and logical ones have to be derived by a sophisticated procedure.
- Design of the document type model (stage 4): On the basis of the catalogue of relevant semantic components (RSCs), a document type model has to be designed. Therefore, all selected components were organised in an XML-typical hierarchy. Depending on the defined target, either structure or contents orientation or a combination can be used. Finally, the document type model has to be implemented, i.e. noted according to XML and transformed in a DTD.

2 Description of the Three Proposed DTDs

The first DTD was developed by the Department of Business Information Systems and Operations Research (BiOR), Chair: Prof. Dr. Heiner Müller-Merbach, University Kaiserslautern. The second DTD was presented by the Institute for Technical and Business Information Systems, Chair: Prof. Dr. Claus Rautenstrauch, Otto-von-Guericke-Universität Magdeburg. The third DTD was proposed by the Institute of Information Systems, Chair: Prof. Oliver Günther, Humboldt University of Berlin.

2.1 DTD from University Kaiserslautern

The target was to develop an integrated DTD for XML-based corporate environmental reports (CERs), suitable for sophisticated customisation. Consequently, requirements of relevant regulations, standards, and guidelines and other academic proposals had to be taken into account. The DTD should be comprehensive, integrated, and flexible in order to serve as a generally accepted framework. As such, the

DTD should stimulate the process of international standardisation, and provide helpful guidance for companies moving towards integrated, internet-based, and customised corporate environmental reporting.

According to the defined target, several resources were analysed for identifying the pool of all PSCs (stage 2). Firstly, relevant regulations and standards (e.g. EMAS II, EN ISO 14031, DIN 33922) as well as common guidelines (e.g. future/IÖW, UNEP) were examined. Thereby, regulations with standard legal basis, other standards, and crucially important guidelines are considered as restrictions. Secondly, in order to take the requirements of target groups explicitly into account, several analyses concerning their preferences in form and contents were included. Especially professional users like investors, funds managers, and financial analysts are going to utilise CERs for decision making processes. According to this general converging trend for integrated reporting, economic and environmental performance need to be disclosed simultaneously. Thirdly, already published realisations of CERs in print media and available on the WWW (Isenmann/Lenz/Müller-Merbach 2001) were considered to guarantee continuousness in reporting practice. Additionally to semantic components, also typical logical components like heading, paragraph, abstract, chart etc. were derived from already published realisations of CERs.

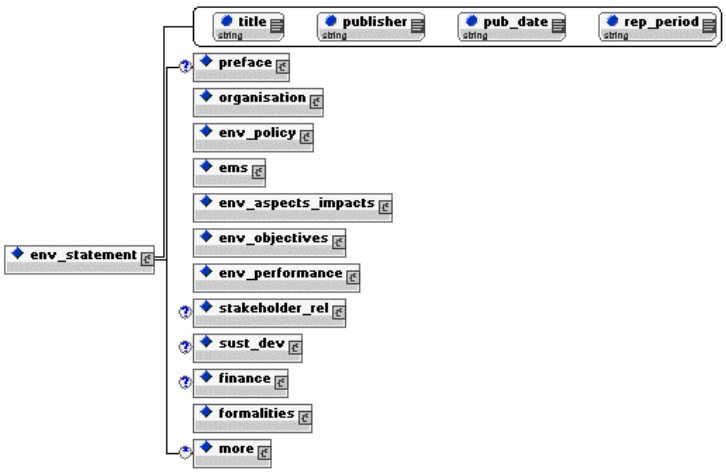

Fig. 1: Top level of Kaiserslautern-DTD (Lenz/Isenmann/Reitz 2001, 63)

Of the pool of possible components, a catalogue of actually relevant semantic and logical ones were derived by a sophisticated procedure (stage 3). Restrictions, target groups' requirements, and document instances were verified for consistency, redundancy, or rejection according to measured relevance. The result was a total of 119 RSCs. Every component was specified whether it must be (required) or might be

(optional) integrated into the CER. In addition to semantic components, the relevant logical components were extracted from published document instances.

On the basis of the catalogue of relevant semantic and logical components, a document type model was designed (stage 4). For that reason, all selected components were organised by an XML-typical hierarchy. In order to exploit the potential of XML-specific benefits, a combination of pure structure orientation and contents orientation was preferred. Subsequently, the document type model was noted in accordance with XML requirements and implemented by an XML-based DTD. The element structure was constructed according to the top-down-principle. First of all, the root element „env_statement“ and the containing elements were defined (fig. 1). Following, step-by-step the child elements and their attributes were defined. CERs prepared in line with this proposed DTD are comparable and can be benchmarked more objectively by machine processing. As a result, such a DTD improves corporate environmental reporting for companies as well as for target groups and other stakeholders involved in the field of corporate environmental reporting.

2.2 DTD from Otto-von-Guericke-Universität Magdeburg

The main goal was to develop a DTD, on which the company Hasseröder Brewery Ltd. can base its environmental reports. Furthermore, designing environmental reports in compliance with fundamental guidelines, relevant regulations and standards should be as flexible as possible. Document instances adapted from this DTD should be reusable as well as generated and processed automatically. Especially for the integration of material and energy balances, impact balances and balance evaluation within the environmental report, adequate semantic components have to be provided. Normally environmental reports will be created in a partial manner by several employees working in different sections or departments rather than by an individual person. Conclusively the different report sections will be composed to an overall report. When designing the DTD, this circumstance has been taken into account.

General conditions for the document model resulted from requirements stated in the EMAS regulations and DIN 33922. Besides, there exist further guidelines from national and international organisations, or branch associations. Extracting standard regulations for structure and content components leads to initial PSCs for the document model. Conducting cross-branch document analyses of environmental reports, considering a core of convergent information needs of important target groups and the described purpose determines further PSCs, such as:

- semantic components for the presentation of material and impact balance, and balance evaluation,
- technical terms and their definitions, whereby environmental reports become clearer and easier to understand,
- time-related words or set of words, e.g. “in the last period” or “in 2000”,

- specific concepts that have to be emphasised in particular,
- meta data containing information about classification and management of documents and their contents.

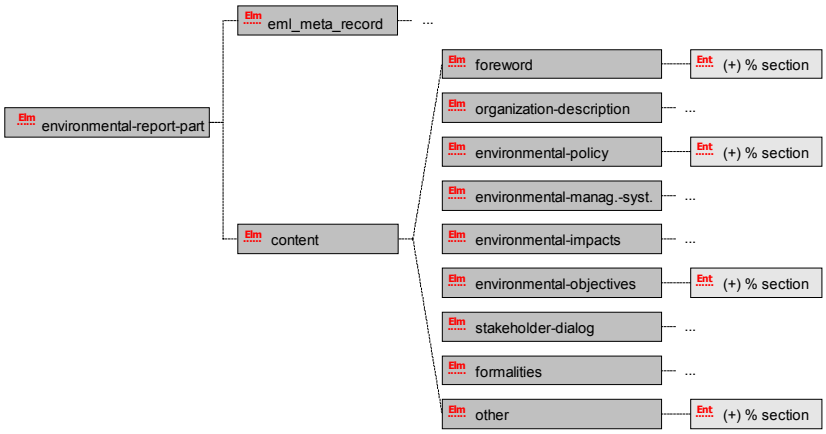

Fig. 2: Top level of Magdeburg-DTD

Those PSCs have been chosen which are required due to their frequency in real documents, due to convergent information demands as well as guide lines for the document model. Remaining PSCs are to be weighed according to their factor allocations and those components exceeding predefined weighting factors have been added to the model. The resulting set of relevant semantic components (RSC) represented the framework for the developed document model (Krüger 2001, 84). Integration of meta data into the environmental report model is based on results of the work group „Environmental Markup Language“ (EML). The suggested set of core EML meta data elements is capable of collecting classification and management information for environmental reports (Arndt et al. 2001, 135). The document type model shows a hierarchical structure in order to allow direct access to document elements and their efficient evaluation. Therefore, type-specific semantic components from the set of RSCs were added to the document hierarchy as structure-supporting elements. Thus, an environmental report based on this document type model simply consists of meta information and links to particular report parts. Each environmental report part represents a special content as shown in fig. 2 and generally consists of a heading and a various number of paragraphs, tables, lists, figures, or sub-segments. The remaining semantic components are used to further concretise the structure elements.

Within the environmental report model the integration of material balances and balance evaluations plays a prominent role, because they are not produced by an author, but are able to be generated automatically. Therefore necessary elements have been structured in a way that environment-relevant data collected by the company can be represented as accurate as possible. It will be distinguished between

input and output balances. Based on that fact materials and energies can be allocated to individual material categories. Finally the document type model forms the basis for the implementation of the DTD in XML.

2.3 DTD from Humboldt University of Berlin

Instead of using a single document for the whole environmental report, we split the environmental report into components by using XML entities, which might be distributed over a network. According to DIN 33922, an environmental report for the public shall include at least the following semantic components (DIN 33922, 1997):

- Basic Information Block: the basic information block consists of:
  - a description of the organisation's activities,
  - a presentation of the organisation's environmental policy and program,
  - a description of the organisation's environmental management system;
- Presentation of Significant Environmental Figures;
- Assessment of all Significant Environmental Issues;
- Declaration of Formal Requirements.

```
<!DOCTYPE environmental_report
[<!ENTITY foreword SYSTEM
                 "http://www.hu-berlin.de/notes/vorwort.eml">
<!ENTITY organizationalActivities SYSTEM
              "http://www.hu-berlin.de/notes/taetigkeit.eml">
<!ENTITY environmentalPolicy SYSTEM
             "http://www.hu-berlin.de/notes/politik_v10.eml">
...
<environmentalReport>
  &foreword;
  &organizationalActivities;
  &environmentalPolicy;
  &environmentalObjectives;
  &environmentalManagementProgram;
  &environmentalManagementSystem;
  &environmentalAspects;
  &environmentalImpacts;
  &environmentalPerformanceEvaluation;
  &otherFactorsRegardingEnvironmentalPerformance;
  &formalStatements;
</environmentalReport>
```

Fig. 3: Integration of semantic components into one environmental report

Based on this four basic semantic components on the highest level an environmental report comprises of eleven semantic components (fig. 3). By declaring each semantic component to be an XML entity, we address the requirement for the support of a component-based approach of environmental reporting. This approach allows each person to work on his part of the report independently and later merge their compo-

nents into the final joint document. This approach requires a new DTD whenever the reference to a component changes. Therefore, and particularly because entities can only be declared in the DTD, new environmental reports, which necessarily contain references to new components, must always refer to new DTDs. We propagate the use of so called "internal DTDs" which are embedded in an EML document. DTDs are then a physical part of the EML-document, they do not have to be stored in a separate file (Arndt/Christ/Günther 2001, 351).

Using entities that constitute an environmental report solves the problem of dealing with multiple copies of one environmental report, thereby omitting redundancies. Whenever one component changes, a new XML document can be created that contains the new link or entity reference. For example, if the old chapter "politik_v10.eml" is outdated and replaced by "politik_v11.eml", a new document containing "politik_v11.eml" as an entity reference can be created. The new EML document can either be stored using a different name or can be supervised by a version control mechanism.

3 Harmonisation of the Three Proposed DTDs

As the development of the three DTDs, also the process of harmonisation might be realised according to the design process of Schraml (1997). Firstly, the elements from the DTDs will be conflated. The result is a new multitude of PSCs for environmental reports. Of this pool the RSCs have to be identified. The following rules for adopting components may be applied:

- Components that are included in all three DTDs,
- components that are include in only a subset of the three DTDs but are required either by standards and regulations, users or because of the frequency in already published realisations of CERs,
- other components that seem relevant for CERs may also be under consideration.

```
<!ELEMENT envStatement (foreword, organisation, envPolicy,
              envManSystem, envAspectsImpacts, envObjectives,
              envPerformance, stakeholderDialog, sustDevelopment,
              financialAspects, formalities, other)>
```

Fig. 4: Proposal for a harmonised DTD for corporate environmental reporting

After having identified the RSCs, a new document model can be developed by defining and refining a hierarchic structure. At this, a decision has to be made if the result should be a rather generalised DTD like from HU Berlin or a more detailed and sophisticated approach like those from Kaiserslautern or Magdeburg, providing a substantial orientation. As a first step, the three DTDs will be harmonised on top

level as shown in fig. 4. Finally, the resulting DTD will be enriched by meta data according to the Environmental Markup Language (EML) Initiative.

4 Conclusion

Reporting companies are confronted with the challenge to rationalise their environmental reporting processes. Further, they are forced to meet different information needs and preferences of target groups addressed (Isenmann/Lenz 2002). An XML-based DTD offers an impressive battery of benefits, firstly for efficient preparation and comfortable administration, and secondly for customised distribution and smart presentation of integrated CERs. On the basis of a standardised XML-based DTD, reporting companies are enabled to provide an integrated and customised corporate environmental reporting system, prepared by machine processing and generated in an efficient and automated manner. The standardisation will furthermore support transparency of and confidence in CER substantially.

References

Arndt, H.-K. et al.: Erster Entwurf einer Dokumenttyp-Definition für die Kernelemente umweltbezogener Metadaten einer Environmental Markup Language (EML). In: Neue Methoden für das Wissensmanagement im Umweltbereich, ed. by K. Tochtermann; W.-F. Riekert. Marburg 2001, pp. 135-142.

Arndt, H.-K.; Christ, M.; Günther, O.: Environmental Reporting in the Public Interest. In: Environmental Information Systems in Industry and Public Administration, ed. by. C. Rautenstrauch; S. Patig. London 2001, pp. 347-354.

Deutsches Institut für Normung e.V. DIN (Hrsg.): DIN 33922 Umweltberichte für die Öffentlichkeit, Berlin, Wien, Zürich 1997.

Isenmann, R.; Lenz, C.; Müller-Merbach, H.: Betriebliche Umweltberichterstattung im Internet. Der aktuelle Stand für Deutschland. In: Praxis der Wirtschaftsinformatik, vol. 38, 2001, no. 218, pp. 97-107.

Isenmann, R.; Lenz, C.: Internet Use for Corporate Environmental Reporting. Current Challenges - Technical Benefits - Practical Guidance. In: Business Strategy and the Environment, vol. 11, 2002, issue 3, pp. 181-202.

Lenz, C.; Isenmann, R.; Reitz, C.: Zielgruppenorientierte Umweltberichterstattung von Unternehmen mit XML. In: Neue Methoden für das Wissensmanagement im Umweltbereich, ed. by K. Tochtermann; W.-F. Riekert. Marburg 2001, pp. 57-69.

Krüger, M.: Konzeption und Entwicklung einer automatisierten Umweltberichterstattung am Beispiel der Hasseröder Brauerei GmbH. Diplomarbeit, Otto-von-Guericke-Universität Magdeburg 2001.

Schraml, T.: Operationalisierung der ökologieorientierten Berichterstattung aus Sicht des Informationsmanagements - Konzeption eines Vorgehensmodells zur formalisierten Explikation von Dokumentstrukturmodellen im Rahmen der Umwelt-Kommunikation von Unternehmen. Dissertation, TU Dresden, 1997.