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

Christian Lenz, Ralf Isenmann[1], Mario Krüger, Jorge Marx Gómez[2]

## Keywords

Automation, Customisation, Document Type Definition (DTD), Environmental Markup Language (EML), eXtensible Markup Language (XML), Harmonisation, Internet, Standardisation

## Abstract

Discussions concerning companies' environmental aspects gained considerable importance and challenged business in recent years. Comprehensible and authentic as well as target group tailored 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 XML. In a more detailed fashion, the contribution is dealing with the harmonisation of two current XML-based document type definitions (DTDs) presented 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, no less importantly, for stimulating the processes towards shaping a unifying markup language in the emerging field of environmental informatics, perhaps in the sense of an Environmental Markup Language (EML). From a practitioners' perspective, such a standardised XML-based DTD contributes to employ Internet technologies and Internet services for corporate environmental reporting purposes productively. Reporting companies and their target groups are enabled to exploit the huge opportunities and media-specific technical benefits efficiently. For example, on the basis of such a standardised XML-based DTD reporting companies are enabled to provide a customised and target group

[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

tailored corporate environmental reporting system, prepared by machine processing and generated in an efficient and automated manner.

The one DTD was proposed by the Department of Business Information Systems and Operations Research (BiOR), Chair: Prof. Dr. Heiner Müller-Merbach, University Kaiserslautern, and the other was presented by the Institute for Technical and Business Information Systems, Chair: Prof. Dr. Claus Rautenstrauch, Otto-von-Guericke-Universität Magdeburg. Both of them were prepared autonomously and they are 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, the DTDs are looking quite similar to 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 going back to Schraml (1997): The overall process-oriented development of the DTD is divided in four stages:

1. Target definition (stage 1): The target was to develop a DTD for XML-based integrated corporate environmental reports (CERs). Consequently, requirements of relevant regulations, standards, and guidelines had to be taken into account. The DTD should be comprehensive, integrating, 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 internet-based corporate environmental reporting.
2. Identification of possible semantic and logical components (stage 2): According to the defined target, several resources were analysed for identifying the pool of all possible semantic components including:

   - relevant regulations, standards, and guidelines (e.g. “EMAS II” of the European Community; “Sustainability Reporting Guidelines” by Global Reporting Initiative, European standard EN ISO 14031 on “Environmental Performance Evaluation”; German standard DIN 33922 on “Environmental Reports for the Public”; guideline “Environmental Reports - Environmental Statements” by future/IÖW; ACCA’s “Guide to Environment and Energy Reporting and Accounting”, and recommendations “Company Environmental Reporting” from UNEP,
   - published realisations of CERs,
   - requirements of target groups addressed as key reporting users.

   Thereby, regulations with standard legal basis, other standards and crucially important guidelines are considered as restrictions. In order to take the requirements of target groups explicitly into account, several analyses concerning their preferences in form and content were included. Additionally to semantic components, also typical logical components like heading, paragraph, abstract, chart, perform-

ance indicators etc. were identified. These were captured and recorded on the basis of an analysis of available published realisations of CERs.

3. Selection of relevant semantic and logical components (stage 3): Of the pool of possible components, a catalogue of actually relevant semantic and logical ones were derived by a sophisticated procedure. Restrictions, document instances, and target groups' requirements were verified for consistency, redundancy, or rejection according to measured relevance. The result was a total of 119 semantic components (fig. 1). Every component was specified by two indicators: The one indicator demonstrates whether the component must be (required) or might be (optional) integrated into the CER while the other one points to the resource on which the component is based on or derived from (standard, guideline, requirement of target group etc.). In addition to semantic components, the relevant logical components were extracted from published document instances.

| Nr. | description | r/o | source |
|---|---|---|---|
| 1 | preface | o | instances |
| 2 | organisation | r | EMAS II, A III, 3.2 |
| 3 | organisation description | o | instances |
| 4 | relationship to parent organisation | o | EMAS II, A III, 3.2 |
| 5 | sites | o | future 6.1, I |
| 6 | number of employees | o | future 6.1, I |
| 7 | turnover | o | future 6.1, I |
| 8 | activities | r | EMAS II, A III, 3.2 |
| 9 | products | r | EMAS II, A III, 3.2 |
| 10 | new ecologically friendly products | o | users |
| 11 | services | r | EMAS II, A III, 3.2 |
| 12 | production processes | o | future 6.1, I |
| 13 | plants | o | DIN 33922, 5.1 |
| 14 | environmental sector trends | o | future 6.1, IV |
| 15 | state of environmental technology | o | users |
| 16 | current topics/projects | o | instances |
| 17 | environmental policy | r | EMAS II, A III, 3.2 |
| 18 | environmental guidelines | o | future 6.1, II |
| ... | ... | ... | ... |

Fig. 1: Catalogue of 119 relevant semantic components, extract

4. Design of the document type model (stage 4): On the basis of the catalogue of relevant semantic and logical components, a document type model was designed. 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 accordingly to XML requirements and implemented by an XML-based DTD. 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 and target groups and other stakeholders involved in the field of corporate environmental reporting.