Designing a Document Type Definition for automated environmental reporting based on XML

Mario Amelung

Eudemonia Solutions AG, Kischstr. 30, 39128 Magdeburg, Germany
Fax: +49 391 543 1837   E-mail: Mario.Amelung@epost.de

Jorge Marx Gómez*

Otto-von-Guericke-Universität Magdeburg,
Faculty of Computer Science, P.O. Box 4120,
39016 Magdeburg, Germany
Fax: +49 391 671 1216   E-mail: gomez@iti.cs.uni-magdeburg.de
*Corresponding author

Abstract: In this paper it will be shown how generation, editing and management of structured contents for environmental reports can be supported. The main focus thereby is an automated transformation of eco-balances into XML form and integration into environmental reports. These already consistently structured contents are the basis for an automated generation of environmental reports for presentation and publication.

Biographical notes: Mario Amelung is a board member at Eudemonia Solutions AG, Magdeburg. He works in the software engineering arena on different projects, especially related to business information systems. Prior to joining Eudemonia Solutions, Mario Amelung studied computer science in economics at the Otto-von-Guericke-Universität Magdeburg and was a Scientific and Teaching Assistant at the Technical University Clausthal. His masters thesis deals with the conception of automated environmental reporting inside the Hasseröder Brewery Ltd. In 2002, Mario Amelung, along with three graduates from the University of Magdeburg, obtained the international award of environmental informatics from the German Informatics Society (GI).

Jorge Marx Gómez studied computer engineering and industrial engineering at the Technische Fachhochschule Berlin. He is a Lecturer and Researcher at the Otto-von-Guericke-Universität Magdeburg, where he also obtained a PhD degree in business information systems with the work 'Computer-based approaches to forecast returns of scrapped products to recycling'. At present he is a Visiting Professor for business informatics at the Technical University of Clausthal. His research interests include business information systems, e-commerce, material flow management systems, life-cycle assessment, eco-balancing, environmental reporting, recycling programme planning, disassembly planning and control, simulation and neuro-fuzzy-systems.

1 Motivation

Environmental reports were published in Germany in the early 1990s for the first time. Since then more and more businesses have realised the strategic competitive advantage resulting from environmental reporting and do not only see it as a measure of public relations anymore. In the same way, both the addressees' interest in and the need for information have steadily increased. Thus, comprehensible and authentic reporting is required. However, businesses are not able to achieve this without having established environmental management and controlling within the company, which enables the gathering and processing of the required data. The Hasseröder Brewery Ltd. tends to self-dependently contribute to reduction and – if possible – prevention of environmental impact produced due to their industrial operation. The brewery's environmental certification after DIN EN ISO 14001 includes building and publishing of an environmental balance that will contain details on their environmental policy and goals as well as their environmental programme and its impact on the business and the region. With two previous projects the Hasseröder Brewery Ltd. established the data basis for a comprehensive and high-quality environmental reporting. Thus, an overall material flow model of the brewing process is already available. Each process step is modelled in detail and underlain with rational operative data [1]. In the context of the second project, a ratio system matching the brewery's needs and data basis was developed, and balance evaluation routines implemented in the past were adapted [2]. Models and estimations are available in consistent form, so that environmental reporting possibly can follow-up on it. Generated environmental reports are necessary to meet high-quality demands, but should also be producible with limited resources. Enterprises that decide on publishing environmental reports are confronted with plenty of informational challenges. One is to gather and evaluate relevant environmental data; the other is target-group oriented information editing, efficient report generation and effective distribution as well as proof of credibility and competence.

This paper focuses on environmental reports in their capacity as documents. Today, in the majority of cases, these reports are only available in layout-oriented specifications on paper or in electronic form. Therefore, environmental reports are neither able to be processed and modified nor transformed into other information products in an automated manner. Thus, the primary goal of this paper is to show how, on the basis of internationally standardised formalisation constructs, environmental reports may be made manageable by machines and therefore can be successfully generated and published. Of particular interest is automatic integration of environmental information available in the business, especially of eco-balance data.

2 Environmental reporting

The term 'environmental reporting' embraces all measures to properly and truthfully inform business-internal and/or -external target groups about environment-relevant issues concerning the enterprise or the regional location. The objective of reporting may be environmental protection goals and activities, description of problems, achievements and environmental effects, which are caused by the enterprise's operation and produced products. Environmental reporting done by enterprises is either voluntary,

or non-voluntary or legally obligated. As voluntary reporting is connected with expenses, it has to financially or intangibly contribute in order to attain environmental goals. Environmental reporting internally provides chief executives with necessary information for environment-oriented management and success control of measures taken so far. Externally, voluntary environmental reporting, e.g. fulfils the function of documenting the enterprise's environmental engagement or serves as an instrument to publicise environmental business efforts [3]. Non-voluntary environmental reporting can be triggered by campaigns of environmentalists or press reports about pollution crimes and failures. Due to pressure by media, politics and customers, the management then has to react with well-directed information and comments [4].

Particular enterprises may be forced to generate environmental reports by legal regulations. Documents that have to be created in this regard are, for instance, emission statements according to Federal Emission Control Law [5], information on security measures defined in Failure Law [6] or waste balance sheets and waste economic concepts under the terms of a Recycling Management and Waste Law [7,8].

Instruments at disposal are the environmental report and the environmental statement. As opposed to environmental statements, the content of environmental reports is not regulated by law [9,10].

In condensed form, for environmental reports and statements following minimum requirements on the context can be stated [11]:

- foreword by management

- description of business operations

- environmental guidelines, programme and goals

- environmental management system

- overview and evaluation of material and energy flows

- important environmental issues regarding products and services

- direct addressing of important recipients

- formal specifications.

Additionally, further emphases can be assigned and integrated in environmental reports such as information on issues of heath protection or involvement and training of employees in environmental questions.

3   Environmental reports based on XML

Machine processing, analysis as well as reuse of documents and parts or contents of them make specific demands on their structure. In order to configure and support the process of environmental reporting most efficiently possible, the logical structure of the existing diversity of environmental documents has to be characterised by means of the mark-up language XML (Extensible Markup Language).

3.1 Document structure

Documents are gathered data concerning a specific topic. This consolidated information is either time-dependent (video, audio) or time-independent (text, image) and organised in logical or graphical structure. Graphically structured documents emphasise on the design of document elements that are visible for the reader. In contrast to this, logical structure relates to the meaning of these elements intended by the author and, for the most part, is designed hierarchically. In most cases logical elements are described by a combination of structure- and content-oriented mark-up characteristics. Therewith, optimal processing, reuse and search for particular data within the document are possible.

3.2 Extensible Markup Language (XML)

With the means provided by XML, documents can be structured logically. Therefore, so-called tags have to be embedded in continuous reference text. Tags either as tag pairs indicate beginning and end of a particular text part or as a single tag, for instance, a line break. A single tag or a tag pair, and the surrounding reference text are referred to as an element. Identical characteristics of all documents within one category are specified in a document type definition (DTD). The DTD determines how elements within an XML document are related to each other and also provide the grammar of the document and its elements. Using XML the main focus is put on DTD design. With DTDs, it is easy for developers to implement software for evaluation and style sheets for presentation without knowing the actual document. Using this, it is possible to generate style sheets for XML documents, which then convert those documents into a wished output format. For XML these are Extensible Stylesheet Language (XSL) and XSL Transformations (XSLT). XSLT transforms XML documents into other XML forms. Using these transformations it is possible to process only a few elements or use them several times. XSL converts XML documents, e.g. into the Portable Document Format (PDF).

3.3 DTD design for environmental reports

Eco-balances at disposal (material and impact balances) must be possible to be integrated. When creating a DTD for environmental reports, in the majority of cases, it will be acted upon the procedure model from SCHRAML [3]. This model is orientated on the classical concept of phases in system development and includes individual phases such as target definition and planning, analysis of essential factors, adjustment and selection of relevant semantic components (RSCs), model design and implementation.

Target definition (phase 1): The purpose of this is to develop a DTD, on which the enterprise can base its environmental reports. Thereby, document instances adapted from the DTD are to be generated and processed automatically and should be reusable. Especially for the integration of material balances (material and energy balances), impact balances and balance evaluation within the environmental report, adequate semantic components have to be provided. Designing environmental reports should, in compliance with fundamental guidelines, be as flexible as possible.

Analysing essential parameters (phase 2): The general conditions for the document model result from requirements stated in the EMAS regulations and DIN 33922 [4]. Besides this, there exist further guidelines required by national and international organisations, or branch associations. Extracting standard regulations for structure and content components leads to initial potential semantic components (PSCs) for the document model [3]. Conducting cross-branch document analyses of environmental reports and considering a core of convergent information demanded by important target groups determines further PSCs, such as:

- semantic components for the presentation of material balance, impact balance and balance evaluation

- technical terms and their definitions, whereby, environmental reports become clearer and easier to understand

- specific concepts that have to be emphasised in particular

- meta data containing information about classification and management of documents and their contents.

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 [12].

Selecting RSCs (phase 3): After having analysed essential parameters, these components out of all PSCs are selected, which are required due to their frequency in real documents, due to convergent information demands as well as guidelines for the document model [3]. Remaining PSCs are to be weighed according to their factor allocations and those components exceeding predefined weighting factors will be added to the model. The resulting set of RSCs represents the framework for the document model to be developed [13].

Model design (phase 4): First task of this modelling phase is the development of a hierarchical structure typical for environmental reports. Direct access and efficient evaluation require precise description of document elements. Therefore, type-specific semantic components are added to the document hierarchy as structure-supporting elements.

Normally, various content modules within an environmental report are produced by employees of different departments at the same time and then composed later. Material and energy balances can easily be generated from this gathered environmental data. Thus, several parts are to be recognised as independent documents, which can be adjusted to a structural framework and can also be dynamically generated. An environmental report, therefore, simply consists of meta information and links to particular report parts. Environmental report parts generally consist of a heading and a number of paragraphs, tables, lists, figures or sub-segments. In addition, individual content modules can be divided into even more sections. Figure 1 illustrates this structure using a tree diagram.

**Figure 1** Hierarchical structure of an environmental report
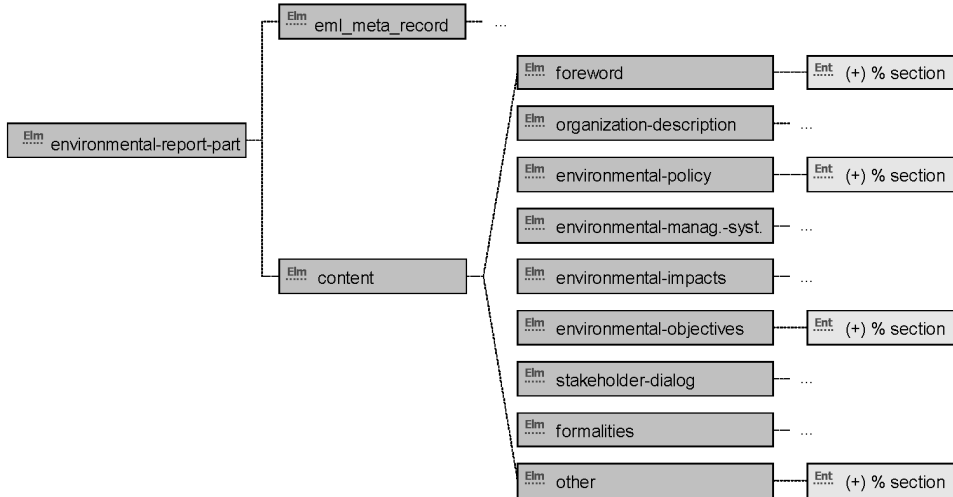
In order to get a better view, self-contained information units are aggregated in the form of entities, as shown in Figure 2.

Figure 2 Aggregation of element types into entities
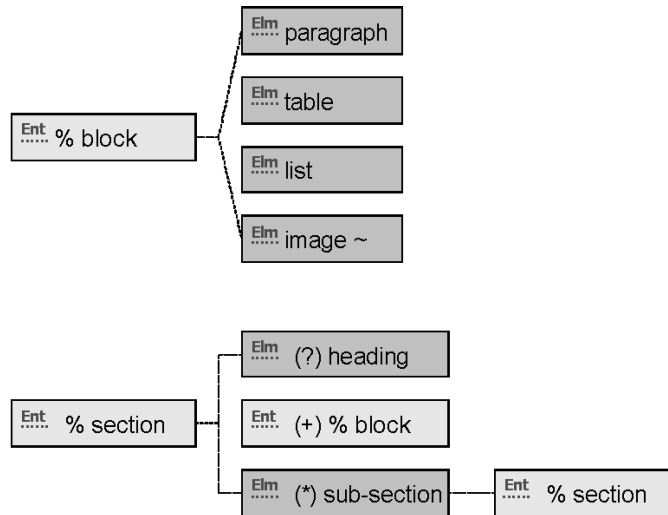
Within the environmental report model, the integration of material balances and balance evaluations play a prominent role, because they are not produced by an author, but are able to be generated automatically. Therefore, necessary elements have to be structured in a way that environment-relevant data collected by the enterprise can be represented as accurate as possible. Figure 3 visualises the developed structure for material balances in the document model.

Figure 3   Structure of material balance in document model
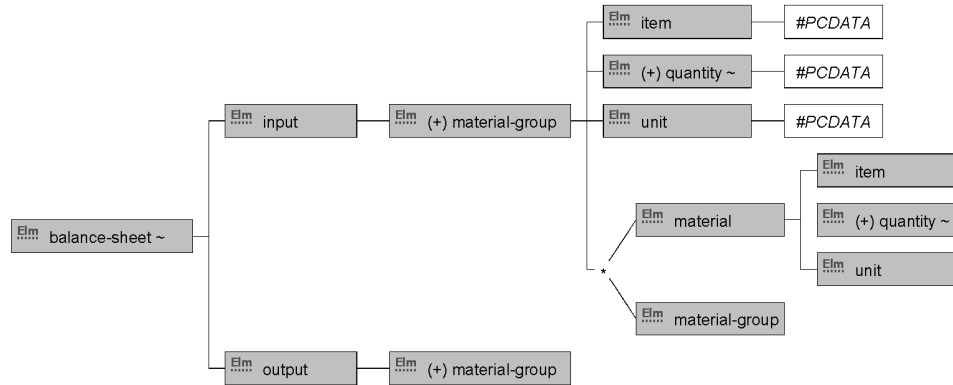
After having modelled the hierarchical document structure, remaining RSCs are used to improve and substantiate structure-supporting elements.

Implementation (phase 5): After having finished phase 4, the model design, semantic components – even if partly not finally tied – are taken into account in the document model. The notation used for modelling basically corresponds to available XML constructs, which has a facilitative effect during the implementation phase.

4   Automatic generation of environmental reports

How can generation, editing and management of structured contents for environmental reports be supported? The consistently structured contents are the basis for an automated generation of environmental reports for presentation and publication.

4.1   Author support

The document basis is the foundation of systematic gathering and management of XML documents and their modules. Various management functions are put on this fundament, e.g. access control, protocol functions, data backup, multi-user capability, administration functionality and query procedures [14]. Creation of these documents requires an editing tool, with which authors can create their texts, mark specific content with tags according to document model specifications and store these files in the central data pool.

4.2   Presenting of information

With the aid of style sheets XML-based environmental reports can be prepared for any desired media and put at disposal according to any preferences target groups might have. The system introduced, therefore, is also equipped with a presentation component, which basically consists of an XML parser and data storage for style sheets, and which has access to XML documents. In case an environmental report is requested, the system checks access authorisation and the document status, associates necessary sub-document, reads current eco-balance data and transforms the XML data into wanted output formats. Besides the output on paper or CD-ROM another opportunity arises: presentation of

environmental reports on the internet. With a connection of a web server to the presentation component, customised environmental reports can be generated and published in accordance with any individual information need that a target group might have [15].

5   Summary and perspective

The objective of this paper was to study whether an automation of environmental reporting is possible and how it can be conducted by enterprises. Applying a process model for document design, developed by SCHRAML, a DTD for environmental reports was written. By the means of shown computer support and automation, the DTD plays an important role in the process of generating environmental reports. A DTD specifies which contents have to exist within the structure of an environmental report and therewith forms a reliable basis for authors. Validity checks will be run on report parts, so that a final document can be easily composed out of them at any time. The document structure, explicitly defined in the DTD, makes it possible to implement application systems, which feature direct access to elements within the document or can fill specific elements with data. Since various DTD design versions for environmental documents have been already developed, it seems reasonable to interchange consolidated findings and hence to support efforts to reach a standard for description and exchange of environmental data in the form of an EML.

References

1   Faustmann, A. and Lutze, V. (1999) *Stoffstrombasierte betriebliche Ökobilanzierung und Bewertung – Fallstudie anhand des Bierherstellungsprozesses bei der Hasseröder Brauerei* GmbH, Master Thesis, Otto-von-Guericke-University, Magdeburg.

2   Wilke, C. (2000) *Analyse, Vergleich und Erweiterung von Kennzahlensystemen und Bewertungsmethoden zur Ökobilanzierung – Fallstudie eines Bierherstellungsprozesses*, Master Thesis, Ottovon-Guericke-University, Magdeburg.

3   Schraml, T. (1997) *Operationalisierung der ökologieorientierten Berichterstattung aus Sicht des Informationsmanagements – Konzeption eines Vorgehensmodells zur formalisierten Explikation von Dokumentstrukturmodellen um Rahmen der Umwelt- Kommunikation von Unternehmen*, Doctoral Thesis, Technical University, Dresden.

4   Clausen, J. and Fichter, K. (1996) *Umweltbericht – Umwelterklärung. Praxis glaubwürdiger Kommunikation von Unternehmen*, München.

5   BImSchG (1991) *Bundesimmissionsschutzgesetz.*

6   StörfallVO (1991) *Störfallverordnung.*

7   KrW-/AbfG (1994) *Kreislaufwirtschafts- und Abfallgesetz.*

8   Rautenstrauch, C. (1999) *Betriebliche Umweltinformationssysteme. Grundlagen, Konzepte und Systeme*, Berlin.

9   DIN (1997) DIN 33922 '*Leitfaden – Umweltberichte für die Öffentlichkeit*', Berlin.

10   EMAS (1993) *Amtsblatt der Europäischen Gemeinschaften Nr. L 168: Verordnung (EWG) Nr. 1836/93 des Rates vom 29. Juni 1993 über die freiwillige Beteiligung gewerblicher Unternehmen an einem Gemeinschaftssystem für das Umweltmanagement und die Umweltprüfung*, Luxemburg.

11 Clausen, J. and Fichter, K. (1995) 'Die Umwelterklärung – der Umweltbericht des EG-Öko-Audit-Gemeinschaftssystems', in Fichter, K. (Ed.): *Die EG-Öko-Verordnung*, München.

12 Arndt, H-K., Freitag, U., Kazakos, W., Schwartz-Glaesker, S. and Westbomke, J. (2001) 'Erster Entwurf einer Dokumenttypdefinition für die Kernelemente umweltbezogener Metadaten einer Environmental Markup Language (EML)', in Tochtermann, K. (Ed.): *Neue Methoden für das Wissensmanagement im Umweltschutz*, Marburg.

13 Krüger, M. (2001) *Konzeption und Entwicklung einer automatisierten Umweltberichterstattung am Beispiel der Hasseröder Brauerei GmbH*, Masters Thesis, Otto-von-Guericke-University, Magdeburg.

14 Rothfuss, G. and Ried, C. (2001) *Content Management mit XML: Grundlagen und Anwendungen*, Berlin.

15 Lenz, C., Isenmann, R. and Reitz, C. (2001) 'Zielgruppenorientierte Umweltberichterstattung von Unternehmen mit XML', in Tochtermann, K. (Ed): *Neue Methoden für das Wissensmanagement im Umweltschutz*, Marburg.