Automated generation of environmental reports – a case study

Jorge Marx Gómez

Otto-von-Guericke-Universität Magdeburg, Faculty of Computer Science, Institute for Technical and Business Information Systems, Department Business Informatics, P.O. Box 4120, 39016 Magdeburg, Germany
Fax: +49-5323-72-7139 E-mail: gomez@iti.cs.uni-magdeburg.de

Abstract: A variety of documents generated within the scope of environmental reporting are based on similar and partly overlapping information. However, in the majority of cases information collected for a particular environmental document cannot be reused for other objectives and if then only with huge additional work input. Data need to be updated on a regular basis, e.g. according to the report period. Therefore computer aided and automated reporting would make good sense. This paper introduces an approach of partly automated environmental reporting. The focus thereby lies on automated conversion of environmental assessment data to XML (Extensible Markup Language) format and integration into environmental reports, which will be shown with the software tool Umberto.

Keywords: content management; ecobalancing; environmental reports; extensible markup language.

Reference to this paper should be made as follows: Marx Gómez, J. (2004) 'Automated generation of environmental reports – a case study', *Int. J. Environment and Sustainable Development*, Vol. 3, No. 1, pp.65–75.

Biographical notes: Dr Jorge Carlos 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. He is a Visiting Professor of 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 Introduction

These days environmental documents – environmental statements and reports in particular – are usually produced in multi-level, more or less structured processes with the participation of several authors and on the basis of various, heterogeneously structured data sources. Defining document structures, the formalisation of environmental reports can possibly help to systemise the work process. If all relevant data exist in a structured form, essential environmental reports can simply be generated by transforming their basic

structures into target structures of the information carriers desired [1]. In this way, for example, diverse ecopolitical company guidelines can be integrated into environmental reports using identifying labels. It will be shown here how creating, processing, and management of structured contents for environmental documents can be facilitated. The main focus lies on automated conversion of environmental assessment data to XML format and integration into environmental reports. The software tool Umberto will be used to illustrate the approach. As a result, integrative and structured content builds the foundation of automated generation of environmental reports for presentation and publication.

2 Document management

The basis of systematic collection and management of XML documents and document components is a central (logical) document base. In a figurative sense the document base is simply a form of data storage including functions such as [read entry], [write entry], [modify entry], and [delete entry].

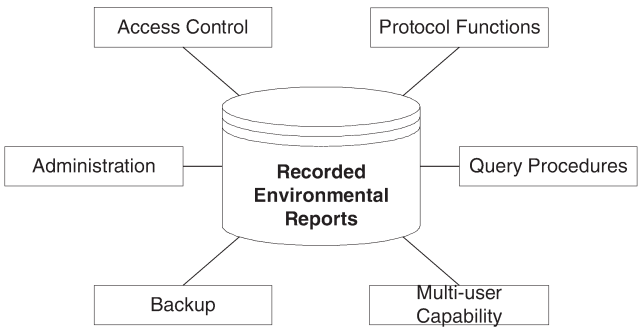
2.1 Document base

A simple document base can be realised with the help of files within a file system. In this case, access control is guaranteed through file system rights. Indeed, for a long time there has been a specific kind of software providing essential document base functions, namely Database Management Systems (DBMS). A modern DBMS is able to store all kinds of objects to databases, including pictures and videos. A possible solution is represented by so-called hybrid forms, which combine a database and file system based external data. Furthermore, the database observes external files, which contain contents that are either too large or too complex, or to which a DBMS can contribute only a little. All structure and management information is located within the DBMS, in order that additional functions can be implemented at a reasonable work input [2].

2.2 Management functions

Management of the document base essentially contains access control, protocol functions, data storage, multi-user capability, management tools, queries, and mass operations as illustrated in Figure 1.

Figure 1 Document management functions



The functions mentioned, the so-called Document Management functions, are supported by Document Management Systems (DMS).

- *Access control:* Access checks are of special importance in multi-user environments. They are based on user management rights and roles. Each user needs to log on to the system, and depending on granted rights, all or only certain interactions are available.
- *Protocol functions:* Protocol functions record modifications in the document base, e.g. whether, when and by whom a document was changed or added. Knowledge produced by protocol functions can be used for protection from unwanted document changes.
- *Data storage:* Effective storage strategies can be implemented more easily if relevant contents are capable of logical, self-contained management and if they are concentrated at only a few physical locations. Furthermore, data-based Document Management Systems can easily be saved by the use of database replication on DBMS level.
- *Multi-user capability:* In order to avoid document conflicts during multi-user sessions, appropriate precautions need to be taken. As soon as the user is in editing mode, the document in use needs to be locked for other users. Due to the fact that DBMSs are multi-user capable and include vital functions, a databased DMS can provide the above-mentioned requirement.
- *Queries:* The more complex the amount of data managed, the more useful are query functions. A single document usually has a determined location within a hierarchy. In general, it is found by navigation through that hierarchy. If, however, only one quality is of specific interest, e.g. all documents with missed deadlines, or documents including ecopolitical means, a system query needs to be generated which requests all information desired.
- *Mass operations:* Mass operations are essential if, for instance, an entire publication needs to be deleted or if all documents assigned to one colleague need to be transferred to another [2].

3 Generation of environmental report parts

In order to generate single parts of environmental reports, XML documents based on a Document Type Definition (DTD) are prepared [3]. Each document represents one part of the report. Environmental reports can either be produced manually or automatically. Report parts build the basis of the entire environmental report.

3.1 Manual generation of XML documents

The generation of documents requires an adequate editor. Authors are able to create documents, label passages in the text according to document model guidelines, and store structured documents in a central data pool. A system that goes beyond document management and facilitates the option to edit contents, is referred to as a Content

Management System (CMS). This editor can either be an external application or an internal editor within the overall system. There are very powerful editors specifically for the generation of XML documents. Indeed, external editors need appropriate interfaces which need to be remotely used through their programme interfaces in order to read, save documents and perform a search in the document base. Another possibility would be to implement a personal editor. Besides editing functions, this editor also performs important operations such as access control, multi-user capability, and queries. In order to adequately support the authors, editors need to be capable of the following:

- providing document templates that already include obligatory elements according to DTD guidelines
- annotating contents with meta data
- setting of XML elements
- insertion of external graphics, pictures, videos, etc.
- linking insertion to other documents or parts of documents in document base
- highlighting XML elements and differentiating from content (syntax highlighting)
- checking XML contents for shapeliness and validity according to the DTD in use
- searching for documents and contents within the document base
- previewing documents, e.g. by the use of transformation into HTML.

The author's environment of CATCH-II [4] is an example for such an application system. The software tool CEdit supports the publisher of health-related documents in creating medical texts, specifying their content by the use of meta data, and introducing both structured mark-up and semantic typecast. For the process of generation of environmental reports workflows can be an adequate means to control the work process. A workflow denotes a part of a business process that consists of sequential and parallel activities. It therefore describes partial processes within the workflow management of a company. Workflow Management Systems (WMS) build the foundation for workflow-orientated application systems. They include functions for the following tasks:

- workflow modelling
- application system process management and supervision
- history logs, with which processes can be reproduced and mistakes detected [5].

From a technical point of view all objects included in workflows represent different states. Interaction between those objects can cause a change in status, which again can trigger secondary processes [2].

Workflow Management Systems can manage the process of generating environmental reports by unlocking specific content modules before previous modules are not completely processed. That way, for instance, the commendation of an input – output balance sheet requires its complete existence, the preface refers to particular contents, and the entire environmental report is not being published before every content module is written. In addition, the WMS considers completion deadlines, e.g. the next company audit, and, if necessary, informs the person in charge.

3.2 Automated generation of XML documents

Essential to all environmental reports is a systematic representation and evaluation of substantial material and energy flows within a company. Necessary data, e.g. raw material and supply flows into the company as well as environmentally relevant pollutants and wastes out of the company, are collected using various methods and, thus, are more or less structured. Now the task is to extract relevant information from the data storage, process and restructure it, and finally integrate it in environmental reports. If the required data is only available in the form of lists, costs need to be incurred in its manual evaluation and conversion to the appropriate report structures. Ideally, the company already has specific software systems for ecobalancing. Eco- and environmental balancing is the generic term for balancing and evaluating consideration of ecologically relevant systems, which can be a product, a production process, or an entire company. The overall goal is to state the main problems of environmental pollution, point out any weak spots, and recommend decisions and actions. Substantial components of eco-balances are:

- objectives and general framework
- property balance (material and energy balance)
- effect evaluation (Estimation of environmental impact)
- balance evaluation.

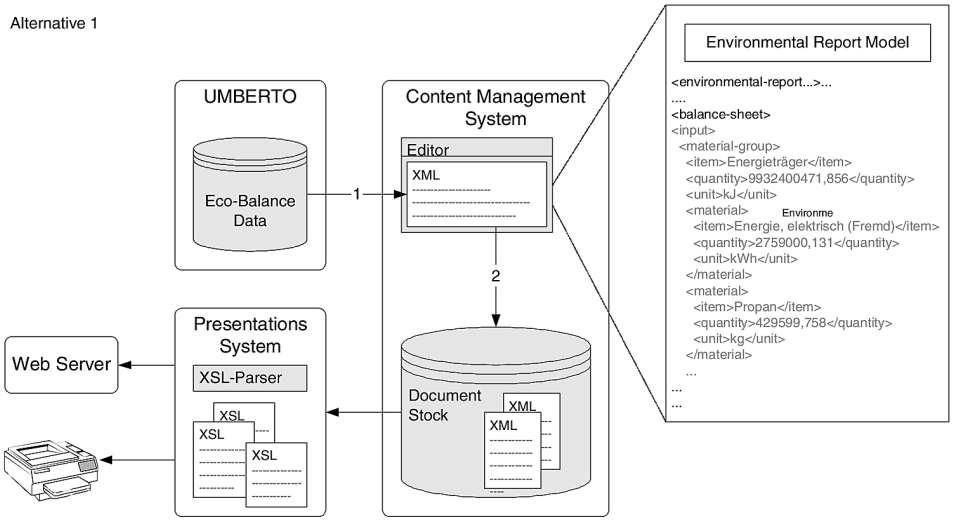
Property balances exclusively represent material and energy flows at the scientific and technical levels, including emissions as input and output factors. They provide necessary information for effect and balance evaluations. Basic elements of a property balance are modelling of the process structure, collecting data, merging process structure and data, and calculating input – output flows. Following this effect evaluations consider the environmental relevance of material and energy flows and other factors. In relation to possible impacts in the range of serious environmental problems, collected data are analysed and evaluated. Here, scientific proof for facts such as global warming, toxicological effects of specific substances, and the contribution of materials to acid rain are added to the balance. For that purpose effect categories and useful indicators are formed. Balance evaluations weight different effect categories to each other. Normally, strong political and social values defining which environmental problems are considered urgent have a lasting effect on eco-balances [6,7]. An eco-balance, therefore, provides all information necessary for environmental reports, from representation of material and energy flows in the form of a property balance to evaluation of environmental effects of products or companies. If eco-balance application systems facilitate external access to the data required, information can be converted to the report's document structure automatically and, thus, they become more efficient.

Thereby, two alternatives arise (see Figure 2):

1 *Option 1:* Property balance and balance evaluation data are extracted, structured according to DTD (1) and stored together with the environmental report resp. parts of the report (2).

2 *Option 2:* A statement is added to the environmental parts of the report (3), according to which property balance and balance evaluation data are requested and inserted (4) at the exact time of transformation to the output format (print or internet page). The benefit is that up-to-date information can be imported to the environmental report, even after eco-balance data has been changed.

Figure 2 Integration of eco-balance data to environmental reports



Umberto is an example of an eco-balance application system. It was developed in cooperation between Institut für Umweltinformatik (ifu) Hamburg GmbH and Institut für Energie- und Umweltforschung (ifeu) Heidelberg GmbH. As a case study, the Hasseröder Brewery eco-balance was generated using the software tool Umberto. For this reason the following software will be introduced, before significant interfaces are addressed in detail and the usage of Umberto for automated generation of environmental reports is specified.

4 Software tool Umberto

The software tool Umberto assists the user in the process of generation of eco-balances. It is applicable independently from the eco-balance methodologies in use and does not assign a particular approach. Umberto is based on the concept of material flow networks. Relevant relationship correlations within an examined system, e.g. a production scenario,

are modelled as a network of interconnected transitions (transformation processes) and stacks (material and energy repositories). Materials and energy of different types and quantities flow along established interconnections between network elements. All material and energy forms are summarised in a material list. Thus, material names can be used consistently and property balance sheets can be compared to each other. The component responsible for property balance output provides various display functions. For selected materials, therewith, numerous standard diagrams (pie charts, bar charts, etc.) can be produced and exported to other applications.

Beyond graphical analysis of eco-balances, Umberto is capable of evaluation with the use of performance measurement systems. Due to the fact that the design of effect analysis and evaluation components is open and flexible, the user is not bound to one methodology in particular. Besides those measurement systems already included in the programme, with the component 'Valuation System Editor' (VSE) further evaluation schemes can be created. Thus, measurement systems following individual needs become possible [7].

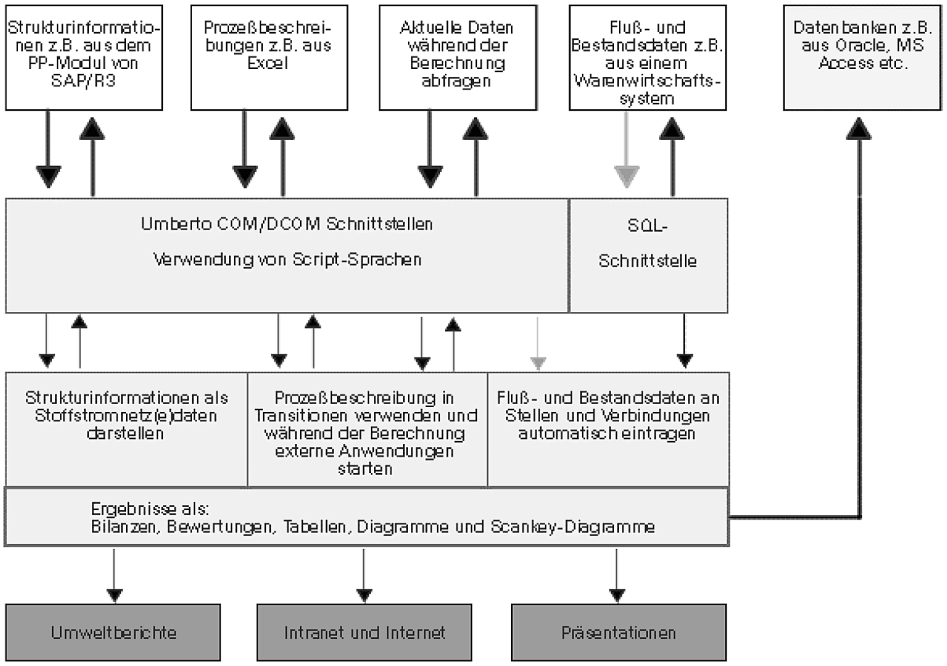
4.1 COM/DCOM interfaces

Umberto, in addition, provides COM and DCOM interfaces. COM stands for Component Object Model and is Microsoft's basic technology for the development of component-based software. Instead of designing an application as one monolithic block, it can be composed from several components. Components represent their functionality in the form of interfaces. Furthermore, distributed COM (DCOM) allows the distribution of components to various network computers. For remote users, the application appears transparent. A COM/DCOM runtime repository guarantees that requests and parameters are securely sent to the target computer and that results find their way back [8,9].

Using COM/DCOM technology, Umberto can link to other programmes locally, over a network, or via the internet (see Figure 3). This enables both customised and time response cost effective usage of the Umberto software tool including:

- personalised import options – once installed updates are simple
- customised input and output masks – each user can view relevant data only
- high automation degree of acquisition of operational data – no multiple data collection
- generation of material flow models – saves time.

Figure 3 COM/DCOM interfaces in Umberto

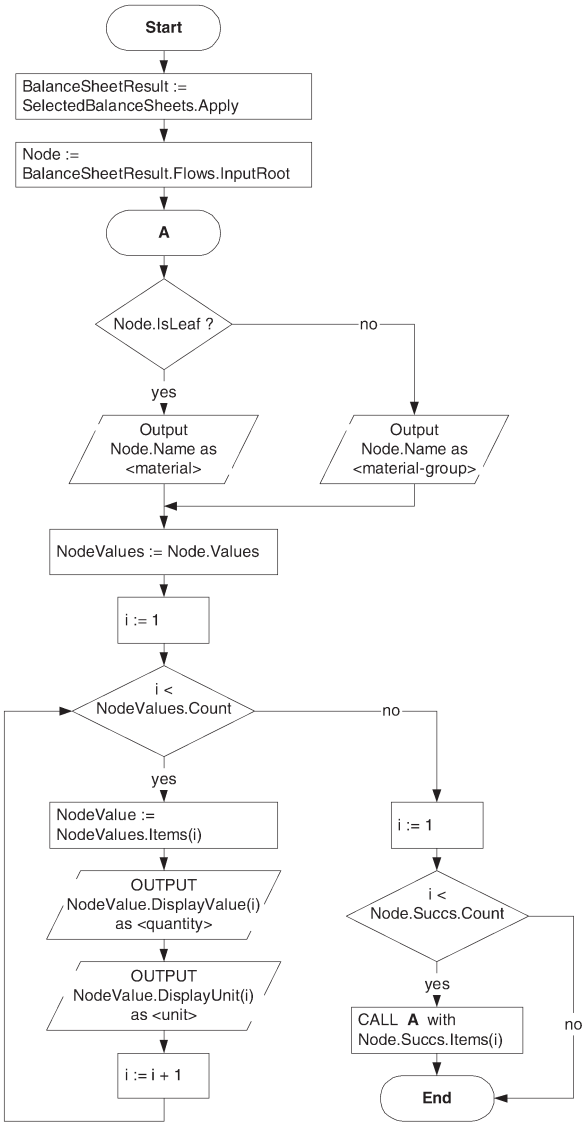


Double input can be avoided due to Umberto's open conception that facilitates a link-up to operational standard software. From a diversity of SAP R/3 modules, for instance, all relevant information can be read out and integrated in the material flow model. If reasonable, data transfer would be possible in the opposite direction as well. With Umberto, optimised system structures can easily be written back to external applications and therefore kept up to date [10].

4.2 Using Umberto's evaluation component

Since Version 4, Umberto provides an evaluation component in the form of a DCOM interface. It allows both performant access to material flow data in other systems and a combination of diverse systems at one consistent user interface or for one specific purpose. Access to the required eco-balance data for conversion to XML is gained through the evaluation component. Umberto itself appears to be idle. However, material flow networks are still modelled using the software tool. The flow chart in Figure 4 shows the process of reading the input side of a property balance out of Umberto's database. Rectangles indicate functions/blocks of functions, rhombuses depict branching points. Parallelograms clarify input and output operations, ovals point out start or end of a programme or function.

Figure 4 Programme flow chart for the readout of eco-balance data



First of all, the net result for the selected property balance needs to be calculated. The necessary data is available in a hierarchical tree structure, so that beginning at the root element all branches need to be processed. Thereby, nodes within the tree symbolise material classes, whereas leaves illustrate single materials. If an entry represents a material, it will be displayed according to customised formats. The output of material classes is analogous. Subsequently, disposable quantities and units related to this entry are displayed. For material classes it might be the case that secondary classes or materials exist within. If so, the algorithm is applied to those entries as well. Umberto's evaluation

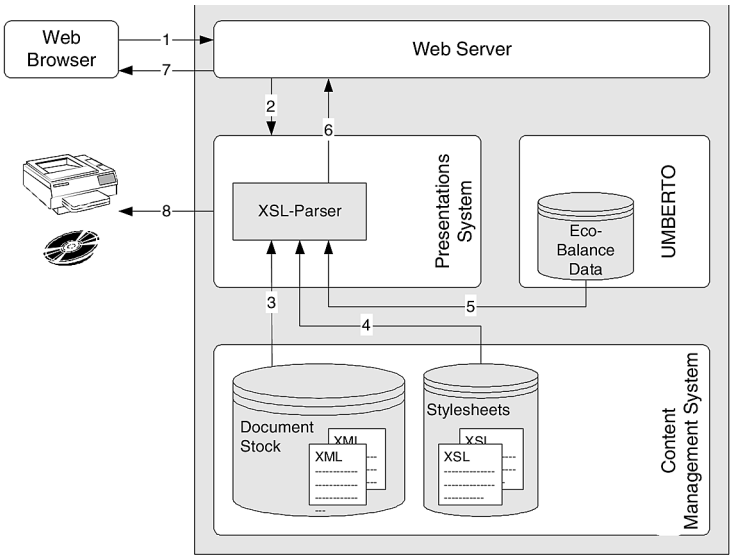
component requires a Windows® computer with Borland Database Engine (BDE) and Umberto base drivers either installed on the same PC or running via the TCP-IP network on another machine. The evaluation tool needs to be registered in the system.

5 Generating and presenting the overall report

Using appropriate XSL style sheets, environmental reports based on XML can be processed for any media and made available according to target group preferences. Thus, a presentation component can be added to the system. This basically consists of an XML parser and a style sheet repository. If an environmental report is requested for a specific medium, the system checks access authorisation and document status, combines the necessary document parts, reads up-to-date eco-balance data, and transforms XML data into the desired output formats. Besides paper or CD-ROM output, presentation of environmental reports in the world-wide web is of significant importance. With a connection to a web server, the presentation component can generate customised environmental reports and view them in a web browser following individual information needs [11].

Figure 5 shows the ideal process of individual requests for environmental reports. Initial generation is triggered by a user query including specific information demands and preferences (1). For evaluations, the web server forwards the user request to the presentation system (2). The presentation component calls the XML document (3) and the XSL style sheet (4) in accordance with transmitted parameters. In addition, necessary eco-balance data need to be read and processed (5). The XML parser applies style sheet instructions to the XML document and generates the environmental report. The output is sent to the web server (6) and finally presented to the user (7). Alternatively, reports can be generated in electronic or paper form and sent to the user by mail (8).

Figure 5 Presentation of environmental reports



6 Summary and outlook

The objective of this paper is to introduce the possibility of automated generation of environmental reports at a corporate level. The focus is on efficient creation and publication of environmental reports as well as automatic integration of eco-balance application systems. The approach described has been used for a case study, whose main focus lay on content design, automated processing, and integration of eco-balance data in environmental reports. Coping with a diversity of addressees and target groups, automated processing of environmental reports seems reasonable and offers full utility potential. Benefits basically arise from structuring and marking document elements, which moreover facilitates future content reuse and adjustment. Based on the common definition of Content Management Systems, integration of eco-balances in environmental reports could be shown with the Umberto software tool. For target group-orientated presentation and publication of environmental reports, the author recommends the integration of a web server in CMS environments.

References

1 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*. Dissertation, Technische Universität. Dresden.

2 Rothfuss, G. and Ried, C.H. (Hrsg.) (2001) *Content Management mit XML: Grundlagen und Anwendungen*, Berlin u.a.

3 Lenz, C., Isenmann, R., Marx Gómez, J., Krüger, M. and Arndt, H.-K. (2002) 'Standardisation of XML-based DTDs for corporate environmental reporting: towards an EML', in W. Pillmann (Ed.) *Proceedings of 16th International Symposium on Environmental Informatics – Environmental Communication in the Information Society*, Vienna.

4 Rösner, D., Dürer, U., Krüger, M. and Neils, S. (2001) 'XML-basierte autorenunterstützung: erstellung und management von multilingualen informationsressourcen', in K. Turowski and K.J. Fellner (Hrsg.) *XML in der Betrieblichen Praxis*, Heidelberg, S., pp.55–66.

5 Rautenstrauch, C. (1997) *Effiziente Gestaltung von Arbeitsplatzsystemen: Konzepte und Methoden des Persönlichen Informationsmanagements*. Bonn.

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

7 Wilke, C. (2000) *Analyse, Vergleich und Erweiterung von Kennzahlensystemen und Bewertungsmethoden zur Ökobilanzierung – Fallstudie eines Bierherstellungsprozesses*. Diplomarbeit, Otto-von-Guericke-Universität. Magdeburg.

8 Microsoft (2000) *Microsoft COM Technologies – Information and Resources for Component Object Model-based Technologies*. http://www.microsoft.com/com/. 18.06.2001.

9 Rosendahl, M. (2001) *Component Object Model*. http://www.uni-koblenz.de/~ros/com/com.htm. 16.06.2001.

10 ifu (2001) *Umberto – Schnittstellen.* http://www.umberto.de/praxis/schnittstellen/. 05.06.2001.

11 Lenz, C., Isenmann, R. and Reitz, C. (2001) 'Zielgruppenorientierte umweltberichterstattung von unternehmen mit XML', in K. Tochtermann and W.-F. Riekert (Hrsg.) *Neue Methoden für das Wissensmanagement im Umweltschutz*, Marburg, S., pp.57–69.