Environmental Reporting of Organizations – Extending the XML-based Standardization Using Topic Maps

Hans-Knud Arndt and Oliver Günther[1]

Abstract

Environmental reporting is an essential component of an organization's communication with its stakeholders. The eXtensible Markup Language (XML), which is platform-independent and particularly well-suited for dealing with semi-structured data, seemed to be a good basis for creating a standardized data format for environmental reporting of organizations. Extending the XML-based standardization we use the new international topic maps standard to describe the complex structure and the meta-information of environmental reporting.

1. Environmental Reporting of Organizations

Environmental reporting of organizations has traditionally been a voluntary method of communicating environmental performance to an organization's stakeholders. Typical objectives of an organizational environmental reporting are (Tochtermann et al. 2001, 333):

- To empower people with information they need to hold organizations accountable;
- To allow organizations and their stakeholders to measure its conformance with its stated environmental policy;
- To identify the significant environmental impacts and their likelihood, severity and frequency;
- To set new environmental objectives and targets.

ccording to the voluntary European Eco Management and Audit Scheme (EMAS) an environmental statement (environmental report) shall be prepared. The international standard ISO 14001 Environmental Management Systems, Specification with guidance for use (ISO 14001 1996) includes establishing processes to report internally and, where desired, externally on environmental activities.

Today environmental reporting of organizations is still dominated by print reports. Even when environmental reports are electronically accessible, they are mostly accessible more or less like paper-based reports, and thus, do not take advantage of the large number of facilities electronic media provide.

A large challenge in this context is to address the target groups. While some groups detailed information require, want other groups short concise environmental information. It is therefore difficult to consider all interests in one environmental report. A containment of the target groups is neither on behalf of the organizations nor the target groups.

Finally environmental communication between the organizations and their target groups is to be regarded on the basis of paper-based environmental reports rather as a monologue on the part of the organizations. Although the paper-based environmental report is a good basis for a dialogue between the organizations and their target groups, it hardly offers possibilities to the addressees for communication. With a paper-based environmental report the communicative devices are limited to the declaration of telephone

[1] Institut für Wirtschaftsinformatik, Humboldt-Universität zu Berlin, Spandauer Str. 1, D-10178 Berlin,
email: {arndt|guenther}@wiwi.hu-berlin.de, Internet: http://www.wiwi.hu-berlin.de/iwi.

numbers and/or post office and/or E-Mail addresses and added reply cards. Communication of the paper-based environmental reporting can be compared to an “one-way street”. Thus, target group-specific communication can hardly be carried out with a paper-based environmental report.

The use of the Internets supports an interactive, hypermedia orientated, differentiated and target group oriented environmental reporting. On the one hand the technical facilities support an efficient production and administration of environmental reports and on the other hand a fast distribution as well as a target-group-specific, interactive presentation are enabled.

In order to exploit the benefits of electronic environmental reporting, the discussion focuses on structuring and standardizing environmental reports of organizations using the eXtensible Markup Language (XML). XML, which is platform-independent and particularly well-suited for dealing with semi-structured data, seemed to be a good basis for creating data formats for environmental reporting (Arndt/ Christ/Günther 2001a, 349). Exchangeability, modularity and quality assurance are the substantial advantages XML offer. A modularity of content takes place, as the content of documents will be overlaid with XML-based meta-information. The meta-information describes the role of document parts and identifies thereby content units. On the basis of these meta-information content units can be accessed, copied, filtered, updated or referenced by other documents. The quality of the documents in form and content is a further important advantage. The once specified structure and it assigned formatting document types remains existing in case of updating.

Using an XML-based markup language for environmental reporting we have the opportunity to handle the structure, content and presentation of an environmental report separately:

- *Structure*: Document type definitions (DTDs) store the structure of environmental reporting. A DTD is a declaration that determines which are the required elements, which are the potential elements of an environmental reporting and in which sequence do the elements have to appear.
- *Content*: XML documents based on a defined DTD store the content of the environmental reporting. An XML document is an instance of a previously defined DTD.
- *Presentation*: Stylesheet languages (e.g. eXtensible Stylesheet Language (XSL) define and process different presentations of one and the same environmental reporting content, such as Portable Document Format (PDF) or Hypertext Markup Language (HTML).

On the basis of XML, a multitude of innovative application possibilities and standardization suggestions were already created. The Environmental Markup Language (EML) is an XML initiative concerning the area of environmental protection. The motivation for the development of an EML is to be seen on the one hand in a uniform basis for discussion in the area of environmental computer science and on the other hand in an XML-based standardized (exchange) format for information and meta-information for environmental applications (Arndt/Günther 2000). The ongoing EML-standardizing process started with the definition of a core environmental metadata set (see figure 1).

```
<!--domainmodell-->
<!ELEMENT    responsibilities (authorOriginator?, distributor,contact?)>
<!ELEMENT    temporal (temporalCoverage?,
             (publicationDate |
             dateOfLastUpdateOfRecord))>
<!ELEMENT    scope (referencePoint?, boundingBox?)>
<!ELEMENT    indexInformation (controlledVocabulary,
             unControlledVocabulary?)>
<!ELEMENT    resourceInformation (languageOfResource, sizeOfResource?,
             typeOfResource,
             formatOfResource?, relatedResources?)>
<!ELEMENT    recordInformation (ID, source, title,
             levelOfMetadata?, abstract?,
             distributionMedium?, accessConstraints?, useConstraints?, lan-
             guageOfRecord?, URL, method?, aggregationLevel?, status?, key-
             words?)>

<!ELEMENT    eml_meta_record (recordInformation,
             resourceInformation, indexInformation?, scope?, temporal, re-
             sponsibilities)>
```

Figure 1
The core environmental metadata set (cutout, Arndt et al. 2001b)

A further goal of EML is to develop further exchange formats for environmental data for different ranges of application (reporting, simulation, material and energy balance, series of measurements ...), which are based on the concept of the core metadata set. A substantial criterion for the acceptance of an EML is the suitability for the respective range of application. Therefore, the EML working group pursues the strategy that for different ranges of application also different data exchange formats are to be considered.

In Germany, there have been at least proposals for XML-based environmental reporting, developed by research groups at the universities of Kaiserslautern, Magdeburg, and Berlin, respectively (Lenz et al. 2002). In order to contribute to the discussion of standardizing environmental reporting of organizations, these three research groups started an ongoing process of an EML-based harmonization.

Two of these approaches and the harmonization approach used single well-structured XML documents for encapsulating the data of environmental reports. These approaches, however, appeared to have three major drawbacks:

1. Putting all the data in one single file makes it difficult for many people to work on a document simultaneously.
2. The integration of a core metadata set into environmental reporting based on results of the EML working group alters environmental report documents and does not enable a strict separation of report content and report context information. Therefore, our group has been following a different approach. Instead of using a single document for the whole environmental report, the report is split into components by using XML entities, which might be distributed over a network. By declaring each part of an environmental report to be an XML entity, we address the requirements for the support of a component-base approach of environmental reporting. Declaring them in another XML document aggregates those entities. The XML document in question is logically one level above the entities. However, its physical location can be anywhere in the network (Arndt/Christ/Günther 2001a, 350).

3. Different target groups need different information extents of the whole environmental reporting. A machine interpretable integration of all different target groups' information requirements in one standardized (hierarchical) XML-DTD is difficult to realize and does not enable a strict separation of report content and report context information as well.

2. The Topic Map Standard

Topic Maps are a new international standard for describing information structures and associating them with information resources. Basic concepts of the ISO/IEC standard 13250 are:

- *Topics*: A topic is defined as an "aggregate of topic characteristics, including zero or more names, occurrences, and roles played in associations with other topics, whose organization principle is a single subject. In the most generic sense, a 'subject' is any thing whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever" (ISO/IEC 13250 2002). Topics are objects within topic maps that represent subjects that their author had in mind when they were created.
- *Topic types*: Topics can be classified according to their type. Classes of topics are defined as topic types by the ISO/IEC standard 13250. Any given topic is an instance of zero or more topic types. Topic types them-selves are topics.
- *Topic names*: Topics have an element form for topic names. Any given topic consists of :
  - a base name (required),
  - display name(s) (optional)
  - sort name(s) (optional, used as sort key).
- *Occurrences*: Topic occurrences are defined as links to information that is indicated as relevant to a given topic, such as articles, monographs, pictures, videos, etc. about the given topic. Occurrences can be online as well as offline resources. According to the ISO/IEC standard 13250, such resources are generally outside of a topic map.
- *Occurrence role*: Occurrences can be classified according to their role. An occurrence role is defined by the ISO/IEC standard 13250 as the "sense in which some set of occurrences is relevant to a topic." The role indicates the way in which the occurrence supplies information to the subject in question (e.g. through being a definition, a figure, an example).
- *Associations*: Topic associations describe the relationships between topics. In the words of the ISO/IEC standard 13250, an association is a "specific relationship among specific topics that is asserted by an association link element."
- *Association types*: Associations can be classified according to their type as well.
- *Scopes*: Topics are characterized of their names, their associations and their occurrences. These three kinds of assignments are defined by the ISO/IEC standard 13250 as topic characteristic assignments. The assignment of a topic characteristic is always done within a specific con-text. The ISO/IEC standard 13250 does not require that this context be specified explicitly. But to do so, each topic characteristic can be provided with one or more context attributes. According to the standard, context attributes are the "extent of the validity of a topic characteristic assignment" and are called scopes. Scopes themselves are topics. Scopes can aid the handling of ambiguity of topics and can be used e.g. for filtering topic maps.
- *Topic maps*: According to the ISO/IEC standard 13250, topic maps are "a standardized notation for interchangeably representing information about structure of information resources used to define top-

ics, and the relationships between topics". Topic maps themselves are navigable without occurrences and any reference to the real world, which is an essential advantage. Relationships between different topics can be illustrated without any altering of information resources in the real world.

opic maps enable multiple, current views of set of information resources. According to ISO/IEC 13250, topic maps can be used (ISO/IEC 13250 2002): "

- To qualify the content and/or data contained in information objects as topics to enable navigation tools such as indexes, cross-references, citation systems, or glossaries.
- To link topics together in such a way as to enable navigation between them. This capability can be used for virtual documents assembly, and for creating thesaurus-like interfaces to corpora, knowledge bases, etc.
- To filter an information set to create views adapted to specific users or purpose. For example, such filtering can aid in the management of multi-lingual documents, management of access modes depending on security criteria, delivery of partial views depending on user profiles and/or knowledge domains, etc.
- To structure unstructured information objects, or to facilitate the creation of topic-orientated user interfaces that provide the effect of merging unstructured information bases with structured ones. The overlay mechanism of topic maps can be considered as a kind of external markup mechanism, in the sense that an arbitrary structure is imposed on the information without altering its original form."

3. Using Topic Maps for Environmental Reports of Organizations

The enabling of multiple, current views of set of information resources is one of the benefits of topic maps. Topic maps can be used to describe the complex structure of an environmental report as well as to describe the context (metadata).

Our entity-based approach is a step towards a semantic-based access to environmental reports using topic maps. On the way to a topic map-based approach we have to go through the following three steps:

1. Using word processing and desktop publishing systems for environ-mental reporting: The analysis of published paper-based environmental report documents gives answers to what primary purposes the documents mainly serve and what kind of requirements have to be taken into account.
2. Using XML for environmental reporting: The strict separation of structure, content and presentation (see figure 2) is one of the major advantages of XML. The development of a standardized DTD gives answers to what are the (required/optional) elements of an environmental reporting and in which sequence do these elements have to appear in an environmental report.

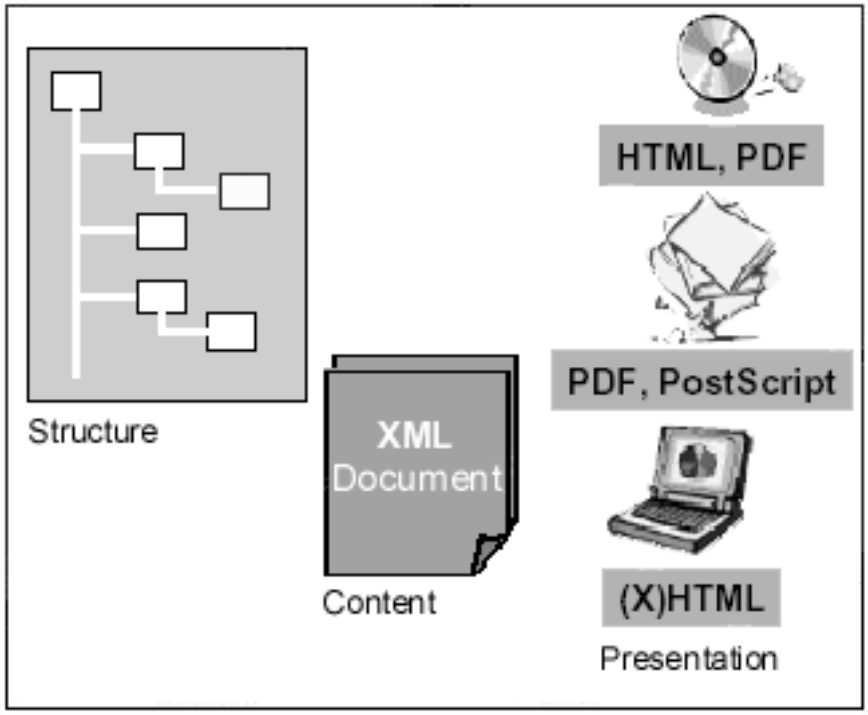

Figure 2
Using XML for environmental reporting
Source: Arndt/Günther 2004, 46

3. Using topic maps for environmental reporting: The enabling of multiple, current views of set of information resources is one of the benefits of topic maps (see figure 3). The development of a topic map for environmental reporting can help to get over the three major drawbacks of the existing XML-based approaches of environmental reporting. Topic maps can be used to describe the complex structure of an environmental report as well as to describe the meta-information.

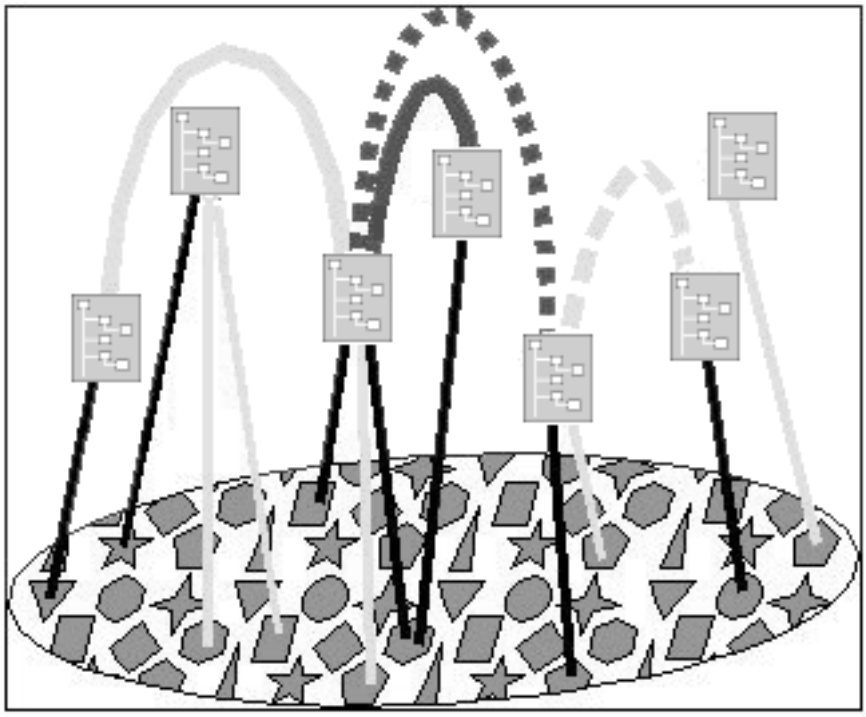

Figure 3
Using topic maps for environmental reporting (from Pepper 2000)
Source: Arndt/Günther 2004, 47

Therefore, we are following a topic map-based approach. Starting off with an entity-based approach to environmental reporting we have to identify parts (entities) of an environmental report in such a granularity that one needs to further decomposition (for the time being). All these identified entities have their own declared structure through an XML-DTD.

The level(s) above these identified entities of the entity-based approach will be replaced by a topic map. In our environmental reporting topic map all identified entities will be represented by topics. The (hierarchical) relationships between the entities will be expressed through topic associations. The use of topic sort names enables to ensure the correct sequence of all identified parts (entities) in the environmental reporting.

Moreover, using topic maps instead of higher leveled entity-based DTD(s) provides the definition of target group specific associations between topics outside the concerning parts of the environmental reporting. Any kind of association can easily be considered, regardless of any specific DTD structure for environ-mental reporting.

In our environmental reporting topic map, meta-information will be represented by topics as well. Using external topics instead of an XML-based meta-data integration we have a strict separation of report content and report context information. The topic map-based approach enables a flexible assignment of meta-information and helps to avoid redundancy within the meta-information of environmental reporting.

An environmental reporting topic map facilitates the representation of any kind of structure and meta-information requirements. It is a semantic web of all declared structure and meta-information requirements concerning an environmental reporting.

So far, in our topic map-based approach we have the ability to represent an environmental report but not the ability to generate different (e.g. target group specific) environmental reports on the basis of the environmental reporting topic map. To do so, we have to obtain (see figure 4):

1. The specific structure of the requested environmental report. Therefore, we mainly use the scope element of the ISO/IEC standard 13250. Specific environmental report structures will be generated dynamically by querying scopes representing structure information and - if necessary - querying other elements (e.g. topic names) of the environmental reporting topic map.
2. The specific content of the requested environment report corresponding to the structure that was generated dynamically. Therefore, beside the generated structure itself we mainly use the scopes representing the context (metadata) of an environmental report. Then, the occurrences of such filtered topics contain the specific content of the environmental report requested.

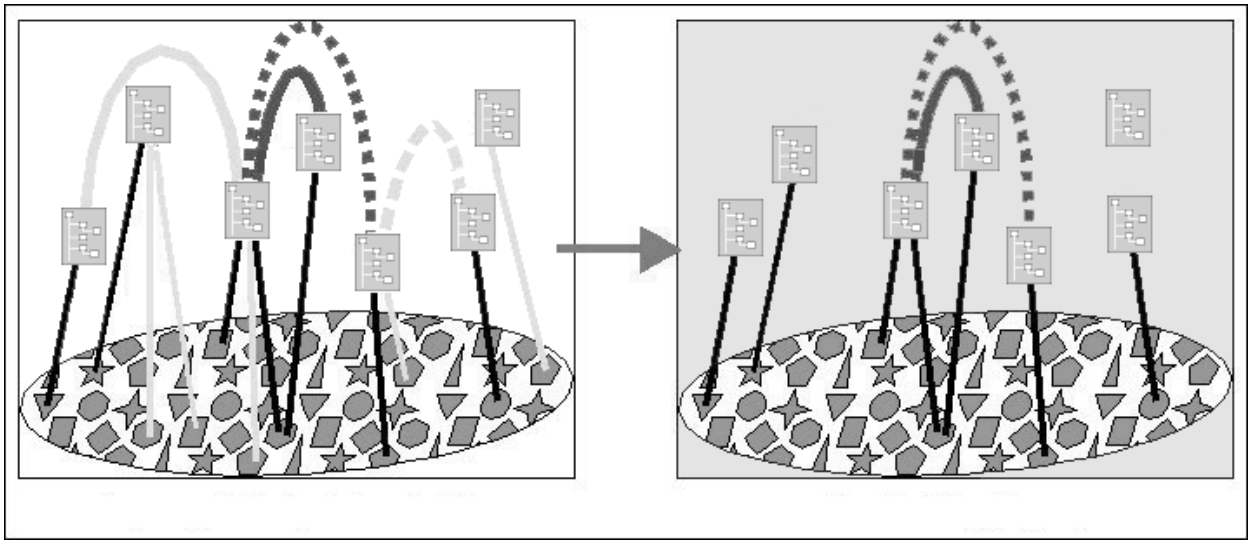

Figure 4
Using topic maps for environmental reporting (from Pepper 2000)
Source: Arndt/Günther 2004, 48

The result of this query process should be designed in a way that each environmental report that is dynamically generated is an XML document with a corresponding DTD. XML documents can easily be transformed to other data formats, such as PDF or HTML using XSL. This transformation enables the multi-media representation (paper, CD-ROM, internet) that is expected from an environmental report.

Instead of handling one or more fixed DTDs for structuring an environmental report, our topic map based approach handles the structure information by spreading it over scopes, topic associations, topic sort names, occurrences (with their optional DTD) and the query statements. Therefore, the query statements represent important knowledge within an environmental reporting and should be handled adequately.

The flexible external handling of structure and context information is one of the major advantages of the topic map approach. Another benefit of this procedure is the enabling of an approach for generating an environmental report more or less automatically. Scoping topics, names and associations, and employing information resources linked by occurrences can be used for a ma-chine-based document assembly of specific environmental reports.

4. Conclusion

We discussed a new approach to environmental reporting of organization that is based on the topic maps standard. The use of topic maps extends the flexibility and the field of possible applications of a pure XML-based environmental reporting. The topic map notation itself can be defined as an XML architecture. While the ISO/IEC standard 13250 uses the Standard Generalized Markup Language (SGML) for the base notation, the members of the TopicMaps.Org Authoring Group provide XML Topic Maps (XTM), an abstract model and XML grammar for interchanging Internet-based topic maps (TopicMaps.Org, 2001).

5. References

Arndt, H.-K., Günther, O. (eds.) (2000): Environmental Markup Language (EML), Workshop 1, Berlin 1999, Marburg

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

Arndt, H.-K., Freitag, U., Kazakos, W., Schwartz-Glaesker, S., Westbomke, J. (2001b): Erster Entwurf einer Dokumenttyp-Definition für die Kernelemente umweltbezogener Metadaten einer Environmental Markup Language (EML), in: Tochtermann, K., Riekert, W.-F. (eds.): Neue Methoden für das Wissensmanagement im Umweltschutz (4. Workshop des GI-Arbeitskreises Hypermedia im Umweltschutz/Workshop 3 der GI-Initiative Environmental Markup Language, Ulm 2001), Marburg, 2001, pp. 135-142

Arndt, H.-K., Günther, O. (2004): Semantic-based access to environmental reports using topic maps, in International Journal of Sustainable Development (IJESD), Vol. 3, No. 1, pp. 34-50

International Organization for Standardization ISO/International Electrotechnical Commission IEC (Eds.) (2002): ISO/IEC 13250 Topic Maps – Information Technology – Document Description and Processing Languages, Second Edition, Genevia (CH), 2002

Lenz, C., Isenmann, R., Marx-Gómez, J., Krüger, M., Arndt, H.-K. (2002): Standardisation of XML-based DTDs for Corporate Environmental Reporting: Towards an EML, in: Pillmann, W., Tochtermann, K. (eds.): Environmental Communication in the Information Society (16th Int. Conference Informatics for Environmental Protection, Vienna 2002), Vienna, 2002, pp. 416-423

Pepper, S. (2000): The TAO of Topic Maps: finding a way in the age of infoglut, Paris, 2002, http://www.gca.org/papers/xmleurope2000/papers/s11-01.html

Tochtermann, K., Keitel, A.; Schütz, T. (2001): Environmental Reporting in Print and Electronic Media, in: Rautenstrauch, C., Patig, S. (eds.): Environmental Information Systems in Industry and Public, London 2001, pp. 333-345

TopicMaps.Org (ed.) (2001): XML Topic Maps (XTM) 1.0 - TopicMaps.Org Specification, 2001, http://www.topicmaps.org/xtm/1.0/