

Ref: 030101

## **SGML PUBLISHING AS A JOINT EFFORT**

**Tuija Sonkkila**

It may take quite a while for a new standard to gain acceptance. In this respect, SGML (Standard Generalized Mark-up Language) has been no exception. SGML was given the status of an ISO standard in 1986. Before and after that, the principal usage of SGML has taken place in the field of technical documentation, where the benefits of getting different end-products for different clientele from the same, structured SGML database have been obvious. It is only now, in the age of multiple new and evolving electronic publishing platforms like the Internet, when SGML is getting foothold in academic circles as well. Indeed the future of SGML looks very promising. This is partly due the fact that the focus is shifting from publishing documents to publishing information.

### **DOCUMENTS ARE CONCEPT OF THE PRINT ERA**

In the age of WYSIWYG desktop editors we are mentally locked to the idea of putting documents together by merely choosing the right fonts and margins. Styles are there for us to get the familiar look and feel of the document. This is only natural, because this is the way the WYSIWYG editors are meant to work. Their aim is to produce individual paper documents. The layout is unseparable from the content which means that the transfer of the document to another software or hardware platform poses a risk, resulting to some loss of properties at best. At worst the document has to be rewritten altogether.

In contrast, if it is the intellectual content of the document we liked to turn our attention to, if we wanted to make sure that whatever the medium in any given time, the message has to be easily transferrable to other formats, then we need a whole different editing approach. This means a separation between the content and the layout; a common understanding that without a proper distinction between the different structural elements of the content there is no life beyond the next generation of publication platforms; a set of tools and policies to accomplish this. In an organisation with a variety of traditional ways of doing things, the change in thinking may take a long time to get through, perhaps never fully. Still, it is worth all the trying.

### **PREDICTABLE STRUCTURE LEADS TO INFORMATION**

In terms of academic publishing, Helsinki University of Technology (HUT) is a major producer in Finland. Annually, it publishes well over 400 titles of research publications in roughly 200 different series, ranging from laboratory notes to Ph.D. thesis. The lifespan of the publications varies considerably, but three things are in common. Firstly, the publishing process is decentralized to the degree that quite often it is the individual researcher herself who takes care of the whole publishing process, from tapping the keyboard to shelving the print run. Secondly,

the only standard that publications are expected to follow concerns layout. Thirdly, the University Library gets a set of copies of every publication.

From the Library's - and ultimately from the end-users - point of view, quality of metadata is the top one concern. Metadata is that part of the publication that identifies it from another, giving a compact, predictable, standardized representation of the origin, content and format of the publication, essential for archival and retrieval purposes. It is no surprise really that HUT publications are quite challenging in this respect, because widely accepted local guidelines of how metadata should be formulated are missing.

There seems to be a relationship between information needs and the structure of information. Research in the field of digital libraries is bringing up evidence that academic users are interested in distinct parts of a scientific document rather than the whole document. As an example, some claim that

*figures reveal what the authors have really done, as opposed to what they wished they had done.*  
1

Given the document is in structured, electronic format like SGML, and given the structure is following a semantically meaningful pattern, SQL-type languages could be used for this kind of sophisticated information retrieval.

In spring 1996, the Library took the initiative of seeking funding from the Ministry of Education in Finland for a four-year project of HUT electronic publishing. An amount of FIM 180K (30KECU) was granted by the Ministry from its special Information Society Fund. After that, a subsequent FIM 130K (25KECU) has been received.

The main goal of the project is to establish a set of procedures for electronic production of HUT serial publication series. Another important goal is to increase local understanding and knowledge about the importance of standards in academic publishing in general, and the benefits of SGML. In-house project partners include the Department of Automation and Systems Technology, the Department of Computer Science and Engineering, and the Computing Centre.

**Q:TOP-DOWN OR BOTTOM-UP? A:BOTTOM-UP.**

Implementing SGML in the publishing process is a tremendous task, a change in publishing paradigm really. Therefore, it is usually thought that for an SGML project to succeed, the involvement and commitment of the whole organisation is needed, from the day one. This might be true in communities with clearly formulated common goals, like companies. But one might argue whether it is foolish to even dream that something like that could ever be achieved in an heterogeneous academic community where, quite understandably of course, individualism is a virtue and departments are traditionally very independent. Other strategies might come more handy. One of them might be called «Request For University Comments», after the well-known procedure of how new Internet standards have to go through an open evaluation process before they are accepted. This is roughly the strategy chosen by the HUT SGML project.

What we need is two or three workable solutions on how to publish successfully using SGML as the underlying concept. By a solution we mean a publishing procedure consisting of the following steps: editing, layout description, creation of metadata, database storage, information retrieval, network delivery, and printing on-demand. Depending on tools and methods used we get a number of different solutions. Regardless of the solution, every part of it has to be truly operable. Otherwise it fails to get attention and approval. Without approval there will be no followers.

At the time of writing this the first prototype done in the HUT SGML project is approaching its completion.

### **THE FIRST PROTOTYPE**

The prototype is based on the assumption that some of the writers might be interested in experimenting with a native SGML editor like FrameMaker+SGML, whereas some others, not being ready to give up their familiar Word desktop editor, would volunteer to act as a fore-runner and use a given template file for later SGML conversion purposes. In that case, the conversion would be done with FrameMaker+SGML, which most probably will be replaced later on by a true SGML conversion tool like Balise. Dublin Core Metadata Element Set, was chosen for metadata. Finally, instead of network delivery in pure SGML format, conversion to HTML was thought to be more appropriate at this stage. Jade, DSSSL engine by James Clark, will be used as the HTML conversion tool. The question of database management is still under discussion.

For those familiar with the issue of whether or not to use an industry-standard document type definition (DTD), it might be mentioned that document analysis resulted in the choice of constructing an own DTD. Future work with subsequent real-life examples will show if this was a wise move or not. SGML analysts tend to emphasize the benefits of industry-standard DTDs (or subsets of them), particularly in network delivery, where stylesheet construction and maintenance may otherwise become a substantial burden.

### **WORKSHOP - STEP TOWARDS REAL LIFE**

A prototype is only a prototype, no matter how technically workable as such. It has to be tested against other types of publications for hints about shortcomings of the DTD. The tools have to be tested by writers for getting feedback. The model of workflow has to be evaluated to find out if it is feasible to put forward at all. This asks for a close cooperation between project workforce and HUT researchers.

Short-term plans of the project include a start-up of a workshop where a small number of HUT researchers are invited to participate. The aim of the group is to bring forward researchers' experience in publishing, and to lay ground for closer cooperation. At the same time, evaluation of the prototype will take place as another sub-project.

### **COOPERATION MAY BE JUST AN EMPTY WORD, BUT IT ONLY NEEDS FILLING**

Cooperation at university level is never a trivial task, partly because of the amount of time and

effort it takes, often without any immediate results. Differences in work culture may be hard obstacles, clashes of interest between organisational units likewise. Nevertheless, cooperation do counts, particularly in publishing, and especially now.

Academic publishing will face fundamental changes in five or ten years to come. Signs are there already. Commercial publishers of scientific journals are losing market, quietly but steadily. At the same time, universities are gradually taking back their former role as academic publishers. To name just few examples, Stanford University's HighWire Press, based at the University's Cecil H. Green Library, announced recently<sup>2</sup> about its work with scolarly publishers in publishing electronic versions of traditional print journals, Lindköping University Electronic Press<sup>3</sup> in Sweden has done innovative work in establishing publishing guidelines, and University of Montreal in Canada will shortly establish an Electronic Press with SGML as the underlying publishing concept<sup>4</sup>. There is still time to learn from their examples. Why not start today?

## REFERENCES

(1) SCHATZ, Bruce et al, Federating diverse collections of scientific literature. *Computer*, 5 (29) 1996, p. 28-36.

(2) EDUPAGE, 15 May 1997. Electronic document, available at:  
<http://www.educom.edu/edupage.old/edupage.97/edupage-05.15.97>

(3) <http://www.ep.liu.se/>

(4) Jean-Claude Guedon in the semi-formal discussion session at the ICC/IFIP Conference on Electronic Publishing '97, 14-16 April 1997, University of Kent at Canterbury, England

---

*Helsinki University of Technology Library,  
Finland  
E-mail: [sonkkila@cc.hut.fi](mailto:sonkkila@cc.hut.fi)*