Review of Track 2

Fytton Rowland
Department of Information and Library Studies
Loughborough, University, UK
e-mail: J.F.Rowland@lboro.ac.uk

When the programme was first conceived, the intention was that track 2 would contain the technical papers, while track 1 would contain those on the socio-economic issues. In the event, more socio-economic than technical papers were submitted, so track 2 contained papers that were not so technical as well as the intended technical ones. This chapter comments on the Monday evening semi-formal session first, before moving on to the sessions in track 2.

The Monday evening session

Paul Evans of Elsevier emphasised the key relationship between a journal and the scholarly community that it serves. He described his attempts to bridge the gap between a scholarly community and an industrial, practitioner community that - in principle - needs the same information. He described how a journal intended to serve both the academic group and the practitioner group was only partly successful -- the scholars read one half of each issue and the practitioners the other half!

Jean-Claude Guédon occupied his favourite slot in the programme -- speaking immediately after Elsevier. Well known for being no friend of the big commercial publishers, he started by stating firmly that scholarly publishing should not be a market, but a project run by researchers for researchers. He distinguished between the creation of authority -- which he defined as publishing -- and the communication between researchers. Published material needs not only to be peer-reviewed, but also to be presented well, with due attention to typographical design. His own electronic journal, *Surface*, has been run successfully for several years, and he now wants to test the economics of scale by setting up an Electronic University Press with a total of 25-30 journals. This would enable the University to test out a number of different pricing strategies. He called for a stable platform, and alleged that the constant flow of changes to HTML arises from commercial interests, not from scholars. In surprising agreement with Paul Evans, he felt that publishers should become facilitators assisting in the development of their disciplines. In a comment which almost fits into track 2, Jean-Claude Guédon also said that he advocated production of SGML production tools and SGML browsers which would be given away to scholars, especially those in the Third World.

Cliff McKnight reported on a project at Loughborough, "Commercial and Free Electronic Journals User Study" (Café Jus). This was a user study of 300 electronic journals by graduate students in a number of disciplines. The students felt electronic journals were easier to access but paper journals easier to use. Drawbacks of electronic journals were network delays, the long-winded routes needed to get to papers from publishers' home pages, and the lack of instructions on how to use Acrobat. The need for user training for electronic journals was very clear.

Jerrold Peterson described a computer-assisted learning project in economics, which incorporated a computer model of the US economy as well as actual teaching material, and assessment quizzes. The text-book market is an important development area for electronic publishing and one which publishers and academics alike should take seriously.

Finally, John Smith described an interesting "thought experiment" in which he had conceived of a different approach to scholarly journal publishing. A "journal" would simply be a web home page with links to quality papers in its subject area, but would not actually carry the papers. These would be submitted by their authors to an authenticating authority -- most likely a learned society -- who would arrange their refereeing. Once accompanied by a seal of approval, the paper would be mounted on the web by its author. Then "journals" -- more properly gateways, or disciplinary resource discovery pages, would point to it. Thus a really good paper could be listed in several "journals". This is an interesting concept which gets away from designing the "horseless carriage" -- a phrase which brings me on to track 2 itself.

Track 2

Two chief themes came through in track 2. One was the "horseless carriage" -- the thought that, at present, by simply converting print journals to parallel form, or by introducing new journals with the characteristics of print journals, we are like the early car manufacturers of 100 years ago, who designed their cars like horse-drawn carriages without the horse. The new media will not look like printed journals. The other main theme that came through in the technical papers was SGML.

Brian Whalley (Queens University Belfast) spoke first of his journal, *Glacial Geology and Geomorphology*. He admitted that the founders were not sure how electronic communication will evolve, but their decided to go ahead anyway (with financial support from John Wiley and Co.). Key concepts were reducing the transaction cost of moving information from person to person, and adding value by the use of the networks. First they undertook a survey of academics in their field and found that what they most wanted was:

1. Peer review -- not a surprise.
2. Rapid publication.
3. Rapid correspondence -- the opportunity to develop a discussion between authors and readers on a short timescale.

Although refereeing is essential, the academics welcome the opportunity of adding links from the journal to other kinds of less academic information. This seems to raise the same question that Paul Evans had mentioned: is it realistic to try to put refereed scholarly material and more practical material in the same journal? Evans and Whalley seem to have had contrasting experiences here. Brian Whalley felt that most of the problems they had anticipated at the outset had gone away as a result of hardware and software advances that had occurred since they began -- justifying their first decision to go ahead without solving all the problems first.

Jim Whitman of the *Journal of Humanitarian Assistance* spoke next. His paper was quite different from the others, and the word "inspiring" is not an exaggeration. The community of people who move in to help when disaster strikes somewhere in the world -- famine, civil war, earthquake -- is, he said, not really a community but a disparate group of people who arrive at

a disaster site having had little communication beforehand. Chaos results. The electronic journal is intended to be an information resource that these people can call upon -- many documents of great value in humanitarian assistance are not well known or well distributed. Their site, "a resource that contains a journal", republishes such documents, many of them very long, points to others mounted on obscure web sites, and publishes new material too. Although housed by, and to a small extent subsidised by, an academic department, the journal is not an academic one. Its purpose is a highly practical one -- to save lives. It must be a resource that is available free of charge, 24 hours a day, anywhere in the world. The Internet is ideal for this.

Simon Polivina of *BIToday* had the difficult task of following Jim Whitman. He too was serving a community that felt it lacked a journal -- the business IT academic community fell between the two stools of management studies and computer studies, the classic interdisciplinary field needing its own journal. Like Jim Whitman, but unlike Brian Whalley, he had had no help from commercial publishers.

In discussion, Jack Rosenfeld raised the question of what role there is for the technical sub-editor. The contrast between the non-technical contributions, which perhaps assumed that content is all-important, and the technical ones, which discussed SGML in depth, was a notable one. SGML is concerned with making publications look right. Another key theme, then, was "content versus presentation".

In the first of the technical papers, Judith Wasteman introduced this enduring theme, of the use of SGML in electronic scholarly journals. She noted that only three of about 60 projects funded by the eLib programme in the UK use SGML at all; the rest use either HTML or PDF. But other major projects outside the UK are using SGML, notably the University of Illinois' Digital Libraries project, and the European Union's ELSA and WebDoc projects. She identified a key issue in the wider acceptance of SGML - choice of DTDs. Despite efforts to achieve standardisation of DTDs for scholarly journal articles, this is not being achieved. Elsevier was again criticised, for standardising on a DTD that is seriously deficient; a well-structured DTD is esential as a standard She ended on a pessimistic note -- SGML is not future proof.

Steve Hitchcock explained the concept, used in the Open Journal project at Southampton University, of separating out the structure of hypertext links from the document itself. The idea is that when a document is added to an open journal system, its text is passed through a processor which adds links on the fly; thus at some future time, it could be input again and create a different structure of links. One application would be to citation indexing, and Freddie Quek of Electronic Press (the proprietors of BioMedNet) described how a paper's links from its own text to its own references could first be recognised, then links from those references to the MedLine database, and then references from MedLine to the other papers concerned, could be introduced by the use of these techniques. Hitchcock himself described a different application in which links from words in a paper to entries in a dictionary could be inserted. The fundamental concept is that the links are held in a separate link server, which is independent of the documents themselves.

This idea, of separating out the content of a document from its structure, is also inherent in the work of Rutledge, van Ossenbruggen *et al.* described in the next technical paper, here in the

context of multimedia documents. They envisage three layers: the *content*, connected by a locating layer which tells the server where all the different objects are to be found; the *structure* of links between these objects, is in turn linked by a style sheet to the *presentation* layer, at the top, which actually presents the multimedia document to the user.

Their application used HyTime, the hypermedia time-based structuring language which is an SGML architecture extending SGML into hypermedia, especially enabling SGML to handle objects which change over time -- such as a video clip of the latest news headlines, for example. In HyTime, as in SGML, there is a problem of differing DTDs, but a new standard DSSSL, converts between them.

Jaime Delgado also spoke about hypermedia documents, but whereas Rutledge was concerned with the structure of the documents, Delgado's topic was the process used for distributed authoring of such items -- a service for the production, management and use of distributed multimedia documents. Multiple editors can work on these items at the same time by the use of a token -- whoever has it can make his or her suggested changes, which are then broadcast to the rest of the group for their approval. Two different standards are available for this, Multipoint Communication Services (MCS) or the more complex Generic Conference Control (GCC). In discussion, there was concern about what happens when different editors disagree, and whether intermediate versions are stored, in case the final version arrived at was not the best one and the editors wanted to backtrack.

The next track 2 session was taken up by two papers from the Technische Universität in München, Germany. Like Rutledge and van Ossenbruggen in the Netherlands, this group works with HyTime and the DSSSL standard for conversion between DTDs. The first paper by Stefan Hermann, "Design by Example", covered an interesting process by which a design for documents in a set could be created by working on just one document in the set. This document was optimally designed by a double-window display, one containing the design tools and the other the current appearance of the document. Once the appearance is optimal, the resulting design specification could be applied to the rest of the set of documents automatically. Different design specifications could be created (for example, one for a screen and one for print), thus staying with Rutledge's and Hitchcock's principle of separating content and presentation.

The other paper from TU München described the MeDoc project, a large project involving several universities and publishing companies in Germany as development partners, and a larger number of universities as guinea-pig users. The idea is to create a secure system for the distribution of electronic publications on a commercial basics. Within the project, there are three major sub-projects:

1. Acquiring content.
2. Developing a secure, billable site for commercial journals.
3. Developing an information-brokering (or selective dissemination of information) service.

Their prototype uses the computer science literature and at the moment, by contract with publishers, they have 50 books and 25 journals in full text in their system. Security is achieved by contact between "user agents" and "provider agents" through the MeDoc intermediary. The topic of security carried through to the next session, which concerned secure management of payment systems for commercial electronic publications.

Daniel Manchala described a system for intellectual property management being developed at Xerox to facilitate commerce in electronic documents. In particular, he discussed an expert system -- the Intelligent Arbitrator Agent -- that could advise in any dispute concerning intellectual property infringements in electronic commerce. In an addendum to his talk, he also mentioned work being done at Xerox to protect text en route to printers from copyright infringement.

Natalie Munard works for a software publisher, Euritis, which has been working on a number of EU projects, and has developed a system, CopySMART, for the management of electronic document commerce. The software tracks transactions concerning distribution of electronic copies and allows delivery only to owners of an appropriate smart card.

The final session of track two presented two interesting applications of SGML-based electronic publishing to humanities scholarship. Ronald Tetreault's project has created electronic texts of Wordsworth's lyrical ballads to facilitate comparison between the different versions. The "vertical" files -- essentially electronic versions of the complete editions of the ballads -- use SGML with the Text Encoding Iniative's DTD. But the multi-window scrolling displays needed for the "horizontal" version -- comparing the same poem across several editions -- have had to be encoded in HTML, because SGML could not provide the functionality.

The BAMBI project, an EU project undertaken in Italy, France and Germany, provides a facility for historic manuscripts to be mapped from the document image to a transcription, and from there to notes and annotations. Double windows allow the user to see the document image and the transcription at the same time. Like Tetreault, this group found SGML unequal to the task, and turned to HyTIME, its hypermedia extension.

A final group, whose paper appears in this book, but who could not attend the conference, also fell into this category of access to historic materials - the Venetian Virtual Archives. Another EU project, it provides access to historic documents about Venice. It is not clear from the paper what technology they have used to achieve this.

Conclusions

So far as the technical papers are concerned, the primacy of SGML and its extension, HyTime, for electronic publishing seems clear. Yet they seem mainly to be appreciated by the academics themselves, and not so much the publishers or even the librarians -- only three eLib projects use it. Commercial publishers, where they do make limited use of SGML, use non-standard or inadequate DTDs. But mostly they are wedded to PDF and Adobe Acrobat, or perhaps its competitor RealPage. The longer term, in which we do not have horseless carriages, seems to lie with separation of three elements:

CONTENT

LINK STRUCTURE

PRESENTATION

which maximises flexibility of linking and of appearance while maintaining the integrity of the basic text. And it would seem that HyTime is a key tool in achieving their objectives. Furthermore, John Smith's new model fits quite well to this concept too: his refereed papers are essentially *content*, and his new discipline-oriented servers are essentially *link servers*.

And where does this leave publishers? This is unclear - Wiley supports David Braisford's *EP-odd* and Brian Whalley's *GGG*, Oxford University Press supports Cliff McKnight's new *Journal of Digital Information*, and yet they do not really know how they are going to make money from pure electronic information.

Fytton Rowland is a lecturer at the Department of Information and Library Studies at Loughborough University, where he is Programme Tutor for the Information & Publishing degree programme. Originally a biochemist, he has previously worked in publishing and information posts with the Biochemical Society, The Royal Society, and The Royal Society of Chemistry.