I'm picking up on Peter Sefton's monster post and one of his phrases suddenly hit me:

> ***academia is one of the few places where PDF is considered acceptable*** *as a means of communication*

I thought about it and I realised – it's true. This awful mess we are in is of our own making. Or rather our own supine acceptance of the PDF served up by scholarly publishers. So why does academia use PDF?

- Because the publishers like it
- Because it looks like a good way to preserve things

And I can't think of any other reason. It's awful to index, to add behaviour to, as a means of developing interoperability (which is trumpeted for repositories but hasn't happened). It is directly against the spirit of the web. HTML has been one of the great successes of the webb (HTTP was another as were URIs).

PDF announces: "I don't care about modern information. I can't think for myself." The medium is the message.

Apart from advertising brochures another area where PDF flourishes is regulatory systems. Pharma companies like PDF because it's much more difficult to search than XML (or HTML) and so harder to find those bits which they don't want found. And regulatory likes it because the pages allow for easy certification .

Is that all academia is about? Helping the publishers certify their page count? And making it difficult to search their pages?

So here's a fuller version of Peter's section:

> ## Scholarly HTML
>
> *Against this background I will confine myself to the dimensions I really care about, which is how to make word processors produce good quality HMTL, and document interoperability. I've been over and over why this is important here, but here's a summary.*
>
> ***On the authoring side, offline word processors like Microsoft Word and OpenOffice.org Writer are probably still the best all round compromised for academic authoring*** *in those disciplines which don't use some other format like LaTeX. For now. I expect this to change soon, we are starting to see document drafting in Google Docs (which lacks citation services and styles and easy embedding of diagrams so far) , and if Google Wave realises its promise then I think it could be an end-to-end scholarly communications platform.*

PMR: Fully agreed. Word processors are complicated because documents are complicated (unless you default to bitmaps such as PDF. I have looked under the cover and PDF is truly awful)

> ***On the delivery side, academia is one of the few places where PDF is considered acceptable*** *as a means of communication whereas on a normal website it is regarded as an impediment to usability. We need to be getting scholarly works into HTML so we can do more with them; meshing them with data and visualisations and delivering them to mobile devises.*

> *While we wait for Google Wave to take over the world, what I'd like to see is a Word toolbar much like the ICE toolbar to support scholarly authoring but with better integration into Word than we have had the resources to make so far here in Toowoomba. It should let people create well structured documents which can be pushed to academic systems; journals, repositories and learning systems and not just in PDF, or Word format, in some kind of formally specified Scholarly HTML. I think that idea had some support at our meeting, but Lee Dirks in particular pointed out that it would need to be done with reference to a stakeholder group who can help define and own this Scholarly HTML thing. I'd be interested in ideas on who these stakeholders might be;*
>
> *Publishers obviously, where MS Research have great contacts.*
>
> *Repository owners particularly the discipline repositories like arXive and Pubmed Central.*
>
> *The eResearch community; I hope that I can get the Australian National Data Service (ANDS) interested in this stuff.*
>
> *The Electronic Thesis and Dissertation (ETD) movement. (My group is involved in this via our CAIRSS repository support service, the Australasian Digital Thesis program in Australia will come to CAIRSS at some point.)*
>
> *The eLearning community, maybe.*
>
> *But actually, where this matters most is on the long tail:*
>
> *Thousands of small repositories and journals are stuck with paper-on-screen because that's all their tools support.*
>
> *The small but growing group of users who want to do more with the versions of their documents they deposit in repositories.*
>
> *I'd appreciate any thoughts about who might be interested in defining a scholarly profile of HTML – a few people told me they're following these posts so please speak up in the comments.*

I'm interested, obviously. My requirements – which Peter knows of course – are that we can embed CML (Chemical Markup Language) and other Markup languages. And that we can start to use RDF (RDFA?).

Please, academia, wake up and embrace the digitalSemantic, not ePaper future.