Scholarly HTML: new approaches to authoring Scientific Papers

I'm in Cambridge, in England. And I have given a talk as part of the leadup to Peter Murray-Rust's Scholarly HTML hackfest.

Friday 11th March1530 Todd-Hamied Room Dept of Chemistry, University of Cambridge

Martin Fenner , Hannover Medical School Cancer Center (DE)

Brian McMahon, Int. Union of Crystallography, Acta Crystallographica

Peter Sefton, Univ. Southern Queensland

Martin, Brian and Peter will present ideas and get feedback in a 2-hour session 1530-1730. The blog posts below give an idea of the material, but the themes are:

- Authors should be able to write articles in the medium that suits them and not have to reformat for every different publisher.
- Authors should be able to include semantic live content (molecules, scripts, tables, etc.)
- The tools should help authors with tedious and trivial tasks such as managing references, cross-referencing, packaging materials for submissions.
- It should be technically trivial to create new types of publication

Peter's brief to me was to talk about whatever I thought was interesting. One of the things that I find very interesting is word processing applications and their relationship, or lack or relationship to the web. I try not to talk about this topic at parties, not that I get invited to many, but at a session entitled "Scholarly HTML: new approaches to authoring scientific papers" I should be able to at least mention Microsoft Word. Right? But more broadly, I'm interested in how we can drag scholarship into the twenty-first century. How can we

In this talk I thought I would talk about the relationship between scholarly communications, computer-based authoring tools and the web.

PMR has done a great job of explaining *why* HTML

HTML? What's that?

Well it's the language that the rest of the world uses to create and publish electronic material – websites, adverts … It's universal. It was designed to communicate science (it happens to sell insurance as well, but science was its motivation). It's easy to author. Even if you don't like pointy brackets there are zillions of free / open tools for creating HTML. So it's obvious that scientists should use HTML for publishing.

In 1995

Working on web publishing tools for a small tech-writing group I built automated scripts for taking styled, structured Word documents and turning them into web pages. I thought:

- "This is cool".
- And "this is really useful".
- And "someone else with more entrepreneurial flair will capture a big market with this.

- And the big one: "The next version of Word will get this right."

In 1997

Working on authoring tools for standards writers at Australia's national standards body I thought:

- "this is cool".
- And "this is really useful".
- And "someone else with more entrepreneurial flair will capture a big market with this.
- And the big one. "The next version of Word will get this right."

Still waiting for decent HTML...

- In 2000 working on authoring tools for an .com servicing online universities.
- In 2005, working on a successful single-source web/print publishing system for the University of Southern Queensland: ICE.
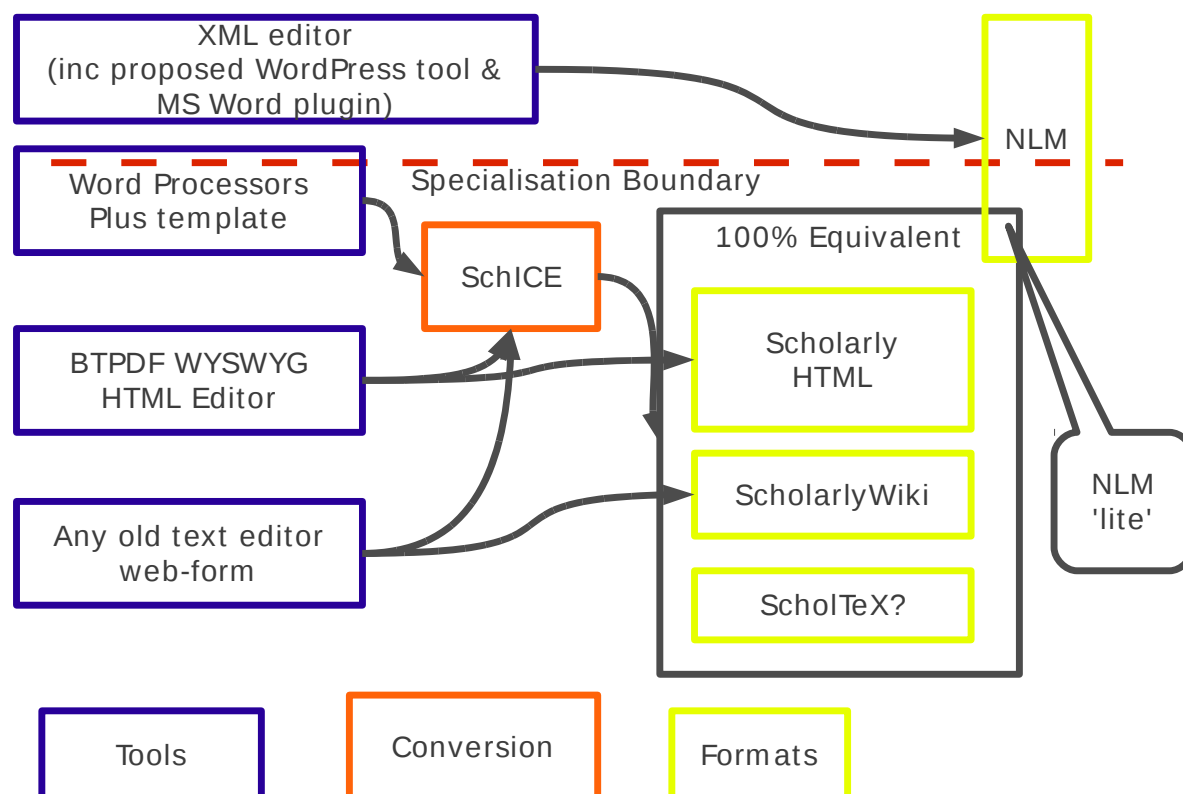- In 2011 as part of the Beyond the PDF movement.

Scholarly HTML

Putting / doing scholarship on the web:

- Documents plus data, provenance etc.
- Packaging / bundling for human an machine use (Epub, ZIP).

Tools and formats

XML editor
(inc proposed WordPress tool &
MS Word plugin)

NLM

Word Processors
Plus template

- - - Specialisation Boundary - - -

SchICE

100% Equivalent

Scholarly
HTML

BTPDF WYSWYG
HTML Editor

ScholarlyWiki

NLM
'lite'

Any old text editor
web-form

ScholTeX?

Tools

Conversion

Formats

Things I'd like to show you

- A usable interface for making structured word processing docs that magically turn into Scholarly
  HTML. (See this post in the Beyond the PDF demonstrator it was automatically redendered from
  this OpenOffice document).

- How we can embed high quality linked-data metadata in docs. (You can have a 'badge' you can
  paste into all your papers that identifies you as an author – like this Peter Sefton).

- A tool for managing research objects of the future – that's papers (or theses or blog post) plus all the
  supporting stuff that goes with them – see for example a chemistry thesis with attached Chemical
  Markup language files.

- How we will be able to embed research data visualisation in a sustainable, easy way. (We're
  working on embedding 3d molecule viewers, having a few problems on the server but it Works For
  Me on my laptop).

Things I hope to show you next week

- Dead-simple packaging of rich research objects into a zip file – ready for upload to a repository,
  journal deposit, etc. We have epub working, log in using OpenId and you can covert that thesis to
  ePub, but ePub is not really suitable for bundling the CML files – we have to work out how to
  handle that.

- Modern, rational citations built around linked data principles rather than where the commas and
  full-stops are.

- Chemistry in WordPress blogs, via Scholarly HTML.That is, linking a document to some chemistry
  in the demonstator authoring enviroment, having it pushed as declarative Scholarly HTML over the

wire to a blog and then having the blog sniff out the chemistry and use a 3d viewer applet to show a molecule.