Real life scenarios for creating and disseminating linked-data publications

In this presentation for Semantic Web Technologies for Libraries and Readers (STLR 2011), which I can't attend in person I want to talk about what happens before things hit the library. I have pre-recorded a couple of demos and asked Jodi Schneider if she would mind introducing the talk for me – I know she has been following this work for a while – maybe she can read out this brief blog post and pretend to be me?

When I submitted this paper I selected three categories for it.

- Strategies for semantic publishing (technical, social, and economic)
- Approaches for consuming semantic representations of digital documents and electronic media and collaboration
- Social semantic approaches for using, publishing, and filtering scholarly objects and personal electronic media

But really it's mainly about the first one, about getting linked data semantics into digital libraries.

The proposal for this paper also had some stuff in it about 'big themes' but given that I am not going to be there in person, and it is only a short demo slot I will not attempt to address those themes.

The ongoing issue with publishing to the web

Ever since the web began to hit the mainstream, there has been a big gap between what's on the web as HTML and the kinds of documents that people write in academia – papers, theses, reports, and so on.

1. For research publications PDF files rule. PDF is what is deposited in institutional repositories, and what people manage (or mismanage) in their personal digital libraries. PDF is not conducive to rich semantics (Yes, it can be done, the web is already ready for linked-data and *all* the action is on the web. And yes, some publishers are doing good things with the web, but they don't typically allow DIY authoring and repository deposit of rich semantic materials).

2. Word processors and tools like LaTeX don't *Just Work* for making web documents it's more like *Just Doesn't Work*.

3. When we start talking about semantics – and wanting to have stuff like RDFa in web pages – it is really hard to do with run-of-the mill scholarship, because our authoring tools don't support formal semantics. (Yes there are XML tool-chains such as TEI – but the heavy-duty XML approach has never been shown to work for large cohorts of non-technical users).

Demos

I want to show a couple of demos

- A method for encoding Linked Data statements in URLS, so they can be used in any system that support simple HTTP hyperlinks. URLs are supported everywhere and will survive being saved as .doc, copied and pasted, emailed and so on, if not a nuclear winter. See the screencast. This includes a lightning look at the The Integrated Content Environment[2] which tries to close the gap between document authoring tools like word processors and the web.

- Packaging all of the above using EPUB, the open ebook standard using various tools and techniques from the Digital Monograph Technical Landscape study using ORE[4] resource maps. See the screencast.

One thing I didn't mention in the screencast is that this work builds on work done by the KnoweldgeBlog project in the UK who are on in the next slot in the workshop – we've never met. Hello!

- I also wanted to look at new models for scholarly objects that allow documents, data and provenance to be managed, reposited and disseminated by demoing The Fascinator Desktop.[3] but unfortunately that's not possible right now as the software in question is being moved to Google Code and the builds are broken – it should be fixed next week. I refer you to a long posting I put together for the Beyond the PDF workshop which has a lot of screenshots.

They key feature of this tool is that it can work with a wide variety of things and bundle them together into a single object,. The idea is to provide a web interface to your hard disk, or Dropbox-like share. In the post I look at a research object consisting of a paper, some data in a spreadsheet, some provenance information, and touch on how semantics about the scientific content of the paper could be marked up using the 'triplink' technique I demonstrated above. (I'll comment below when I am able to post a screencast).

Sorry I couldn't make it to Ottawa, hope you all enjoy the workshop.