Scholarly Publishing using the Integrated Content Environment

2008-12-11

Peter Sefton

University of Southern Queensland

sefton@usq.edu.au

1 Orientation

This is another blog post based on a presentation. What I've been doing lately is working on each presentation as a document with embedded slides before I give it. I think that working this way makes for a more coherent delivery and I can take a bit of time to edit the thing into something a bit more useful than a PowerPoint presentation*. Of course it's only ever half done by the time I give the talk so I have to come back and finish it off and edit to make it reflect what I think I actually said.

Here's the notes for my talk at the recent Open Access Publishing workshop run by APSR. The website bills it as “A two-day Public Knowledge Project Workshop”, while the program says it was a “a PKP User Group Workshop”, either way the idea was to talk about issues around the mission of and the software produced by the PKP project. Website says:

> The Public Knowledge Project is a research and development initiative directed toward improving the scholarly and public quality of academic research through the development of innovative online publishing and knowledge-sharing environments. More...

This was the last APSR event and I'd like to take the opportunity to say thanks to the APSR crew, particularly Margaret Henty for all the APSR events of various kinds over the last few years. These events got us started in the repository business, APSR events helped bootstrap the RUBRIC project.

I started with this abstract:

> **Abstract:** The Integrated Content Environment (ICE) is an open source software application and service-platform for word processor based academic authoring, originally built to support the University of Southern Queensland's flexible-delivery courseware but now expanding into more general scholarly publishing.
>
> We will show a myriad of ways that ICE can be used in the academy. Starting from document creation, ICE uses generic word processing templates which capture the structure and semantics of scholarly documents. The system manages collaborative works in progress using a distributed version-controlled repository, and can publish works to HTML, PDF and domain-specific XML schemas. It has been integrated with several other systems, including the Moodle Learning Management System and DSpace, ePrints and Fedora repositories.
>
> Specific examples will include an journal which uses the ICE publishing system, demonstrations of semantic-web publishing for theses (an outcome of the JISC-TheOREM ICE project) and institutional repository integration.

* As I always say a used PowerPoint (particularly without speaker notes) is about as much use as a used condom. Fun for those who were there, maybe, but lacking appeal for those who were not.

2 Introduction

The Integrated Content Environment (ICE) is an open source software application and service-platform for word processor based academic authoring, originally built to support the University of Southern Queensland's (USQ's) flexible-delivery courseware but now expanding into more general scholarly publishing.

ICE is a product of a software development team now located in the Australian Digital Futures Institute (ADFI) at USQ. The Institute is involved in a research and development projects in eLearning and eResearch.

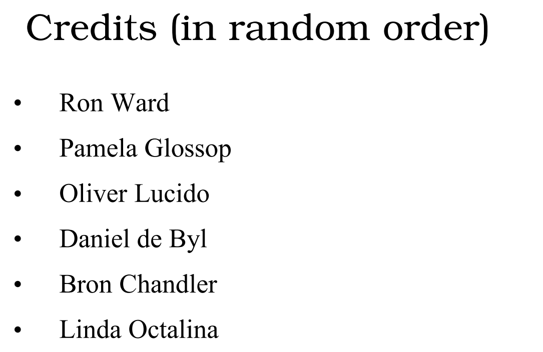
When I first wrote up this talk, I thought it might come out a bit negative. I had this paragraph:

> This presentation I will show-off ICE and talk about how some of its modules can be used in and mashed-up with other systems. Unfortunately, the conclusion to this presentation is not "and they all lived happily ever after". After showing some of what ICE can do, a lot of which I believe is very important to the scholarly enterprise, I will go on to talk about the challenges that lie ahead. All we have to do is change the entire scholarly publishing model and along the way make our publishing systems much more flexible and far reaching.

But actually, the tone of the meeting made it a bit hard to summon the doom and gloom. We had a very productive exchange on how we might mesh our technologies and experience with the PKP systems.

3 About ICE

I kicked off with some credits. ICE is a team effort, I make the most noise but I don't do the most work.



Credits (in random order)

- Ron Ward
- Pamela Glossop
- Oliver Lucido
- Daniel de Byl
- Bron Chandler
- Linda Octalina

ICE is more than one thing. See the website.

ICE components

- Word processing templates (MS Word and OpenOffice.org)

  Easy to use, generic, extensible structured editing.
- A web application for converting word processing documents into HTML and PDF, packing and disseminating them.

  (This normally runs on the desktop – one copy per user)
- A set of APIs so that ICE components can be used in other systems.

The templates are at the heart of the ICE system. They provide a simple generic way to structure word processing documents. It is well known, if not a scientifically documented fact, that using a word processor to create web pages typically results in, shall we say sub-optimal web sites.

ICE is style-driven, not as in fashion conscious, but as in styles-as-named-bundles of-formatting which have structural or semantic significance.

ICE is style driven

About the only styles your word processor can export to a web page are Heading 1 and family.

So ICE defines styles for:

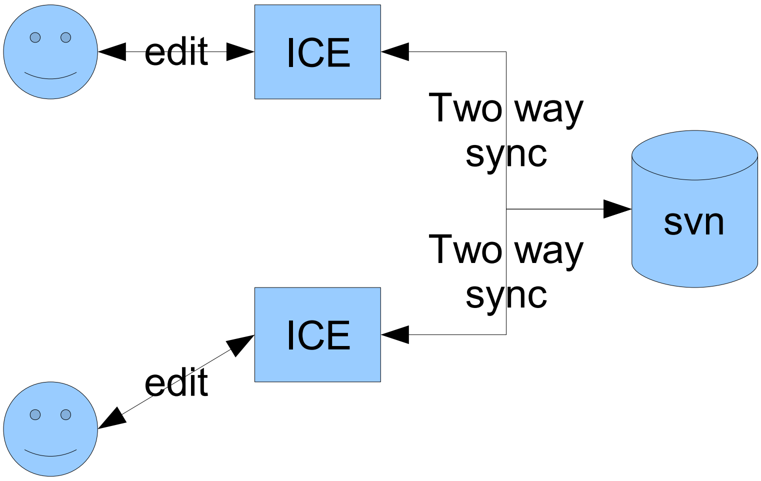
1. **Headings** – with and without outline numbering.
2. **Lists** – lots of flavours including bullets and definition lists and various was of counting enumerations.
3. **Blockquotes** and **preformatted** text.

To convert word processing documents to HTML and PDF ICE uses OpenOffice.org. ICE converts Word documents to the OpenDocument Format behind the scenes, a step which is not necessary with OOo Writer. And from there there's code to turn ODF into HTML. ICE uses Writer to generate web-ready images from the various graphic formats, equations etc.

ICE can be used in lots of ways. It started life as a courseware authoring system. At USQ we have a decades-long tradition of producing printed distance education materials, and for the last decade or so of making the same content available on the web, with added webbish goodness.

ICE for courseware

http://ocw.usq.edu.au/course/view.php?id=13



Starting from document creation, ICE uses generic word processing templates which capture the structure and semantics of scholarly documents. The system manages collaborative works in progress using a distributed version-controlled repository, and can publish works to HTML, PDF and domain-specific XML schemas. It has been integrated with several other systems, including the Moodle Learning Management System and DSpace, ePrints and Fedora repositories.

ICE for theses

ICE can manage book-length content like theses quite well:

- Creates HTML and PDF
- Annotations for supervisor comments
- Integrated data visualizations
- Automated repository deposit

While ICE can be used to manage thesis drafting it doesn't really manage the other processes that go on, Particularlty the examination process where PKP's Open Journal Systems (OJS) might help – OJS is used at USQ to manage the thesis review process for *some* theses.

Speaking of journal systems I showed an example of journal which was published using the ICE publishing system:

A Journal

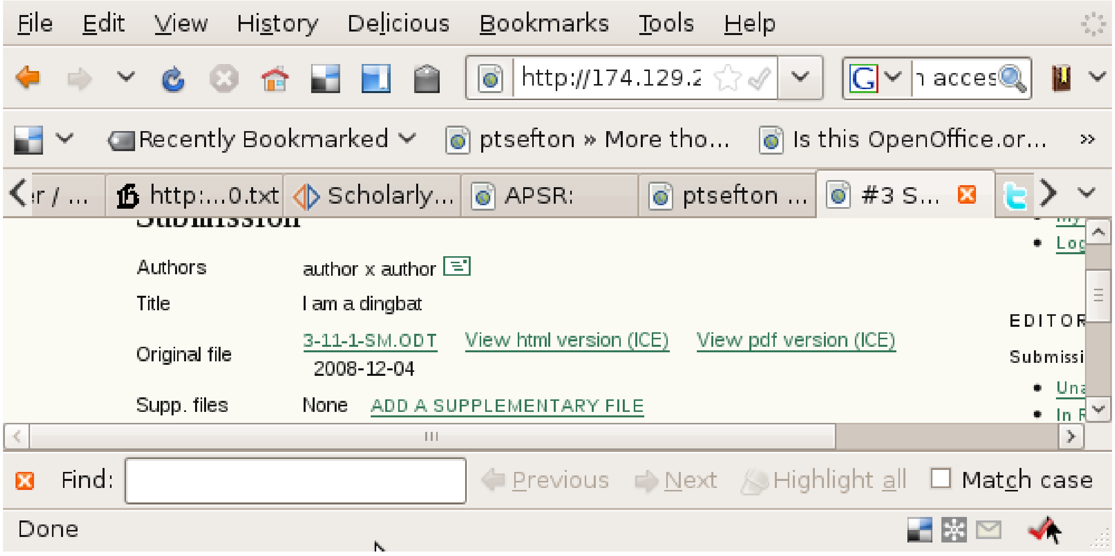
http://www.usq.edu.au/electpub/e-jist/docs/vol10_no1/default.htm

ICE gave no help with the journal workflow, but it did create HTML and PDF renditions of all articles

There's very little overlap between ICE and OJS.

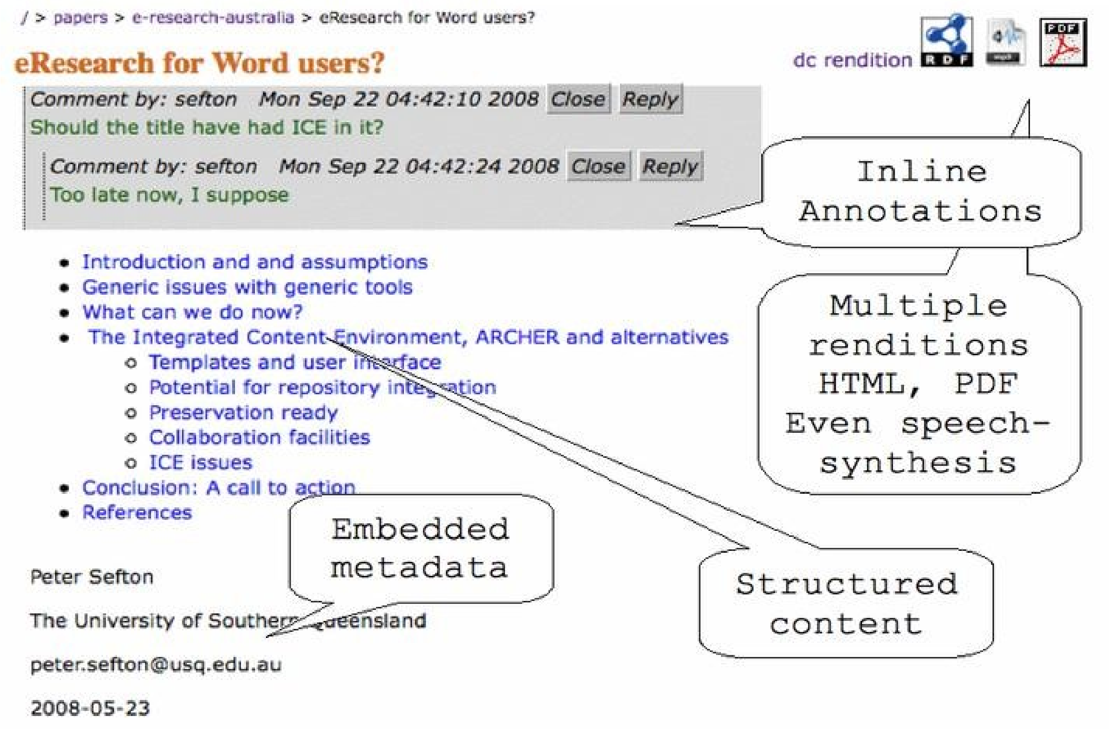
ICE wrt OJS

I was able to demonstrate some work that Linda Octalina has been doing to integrate ICE conversion services into OJS. She's added a basic feature where you can upload an ICE document and instead of treating is as a monolithic word processing file, OJS fires it off to ICE to be converted into other formats, and then shows them to you:



Authors	author x author
Title	I am a dingbat
Original file	3-11-1-SM.ODT View html version (ICE) View pdf version (ICE) 2008-12-04
Supp. files	None ADD A SUPPLEMENTARY FILE

In the above screenshot you can see how OJS now has an HTML and PDF rendition for each file, If we could make this work then OJS would be able to use ICE features like embedding data such as chemistry or inline annotations. Here's a screenshot that covers a few ICE features:



/ > papers > e-research-australia > eResearch for Word users?

dc rendition

eResearch for Word users?

Comment by: sefton Mon Sep 22 04:42:10 2008 Close Reply
Should the title have had ICE in it?

Comment by: sefton Mon Sep 22 04:42:24 2008 Close Reply
Too late now, I suppose

- Introduction and and assumptions
- Generic issues with generic tools
- What can we do now?
- The Integrated Content Environment, ARCHER and alternatives
  - Templates and user interface
  - Potential for repository integration
  - Preservation ready
  - Collaboration facilities
  - ICE issues
- Conclusion: A call to action
- References

Peter Sefton

The University of Southern Queensland

peter.sefton@usq.edu.au

2008-05-23

Inline Annotations

Multiple renditions HTML, PDF Even speech-synthesis

Embedded metadata

Structured content

ICE lessons

What would these mean for the PKP tools?

- **Feedback** early and often.
- We need '**bundle aware**' content stores so we had to build one. (Disseminators, Fedora Commons style are another option)
- People **can and will** structure their documents if there is a good reason for them to do so.
- Learning structured word processing is a **key skill** for academics – why are we not teaching them how?
- We have done the work to define a **generic structured document profile on top of ODT** (didn't realise that's what we were doing!)

The final point above, about a generic structured document profile seemed to resonate with MJ Suhonos from PKP. We've corresponded for a while about getting ICE and his Lemon8-XML project better aligned. Now that we've met s and had a couple of glasses of wine that should be a bit easier.

If MJ decides to use ODF we'll be there to help with the process of getting the Open Document Format word processing format (ODT) plugged into Lemon8-XML as a back-end, a byproduct of which will be what I was hoping for in my previous post, a way to bring unstructured documents into the ICE fold.