WordPress

Introduction

So far in the jiscPUB project I have been looking at word processing applications and EPUB, as well as how repositories and other web applications might support EPUB document production. One of the tasks in workpackage 3 is to look at WordPress as an example of an online tool that's being used quite a bit in academia for both writing and publishing.

> The three main use cases identified in the current plan, and a fourth proposed one: [numbering added for this post]
>
>   a.   Postgrad serializing PhD (or conference paper etc) for mobile devices
>
>   b.   Retiring academic publishing their 'best-of' research (books)
>
>   c.   Present final report as epub
>
>   d.   Publish course materials as an eBook (Proposed extra use-case proposed by Sefton)
>
> http://jiscpub.blogs.edina.ac.uk/2011/03/03/workpackage-3/

The next few posts will explore web based authoring and publishing with a focus on WordPress, and how they relate to packaging content as electronic books.

WordPress can be used in a number of different ways. For this project I am thinking of it as:

- A publishing platform.
- A collaboration platform.
- A content aggregation platform.
- An authoring environment where people might write academic content. (I put this last, because I think it's the most controversial).

All of these overlap, and the same installation of WP might be doing all or none, as might other content management systems being used in academia.

In future posts I'm going to look at building ebooks via aggregation, using the Anthologize plugin, look at an alternative way of building EPUB books from lists of WordPress posts using Calibre, and take a look at Martin Fenner's EPUB plugin for WordPress. In this post I will look at some of the issues around WordPress as used in a couple of projects related to this one, looking particularly at JISC-funded or JISC-friendly work. This is not a survey of how WordPress is being used in academia everywhere – there's no time for that. Please use the comments below if I've missed something that's important to this project.

At the moment, I am thinking that the most compelling match up between the use cases for this project and what is being done with WordPress are these:

- **b: Retiring academic publishing their 'best-of' research (books)**: not so much books but using a tool like Anthologize to draw together papers or other documents.
- **d: Publish course materials as an eBook** (Proposed extra use-case proposed by Sefton): I see great potential for tools like Anthologize as a way of compiling reading packages from web resources and packaging them to take-away on mobile devices, likewise for conference proceedings and programs and other aggregated documents.

And possibly, where people are using JiscPress this use-case: **c: Present final report as epub**.

Publishing platform

A great example of using a blogging platform for scholarship is the KnowledgeBlog project:

> We are investigating a new, light-weight way of publishing scientific, academic and technical knowledge on the web. Currently, Knowledge Blog is being funded by a JISC grant.

And the sites it has under its wing.

- Ontogenesis
- Process
- Taverna

KnowledgeBlogs use the WordPress platform to publish articles and for article review and serves as a live example of a new mode of scholarship. It's a publisher, but not as we know it.

A new entrant in the WordPress backed publishing space (and in the Authoring space) is Annotum which has not released any code, but has very lofty ambitions. I'll come back to Annotum below.

An aggregation platform – bringing together content from elsewhere.

I'll cover this in my next post, looking at Anthologize, which is a promising but immature tool for pulling together stuff from multiple sources and/or authoring it locally, then grouping it with a customized table of contents and publishing to a variety of media.

An authoring platform

I has to be said that WordPress as an editor gets some bad press from time to time. Phillip Lord at KnowledgeBlog advises against using it for authoring.
WordPress is not an authoring environment

> http://www.knowledgeblog.org is hosted using WordPress. It's a very good tool in many ways, but it was intended for and is most suited for use as a publishing tool; most blogs are written by single authors who wish to place their thoughts on the web either for authors or themselves to be able to read. It is not an authoring tool, however. It does not provide a particularly rich environment for editing, and particularly not for collaborative editing. Most people get tired of the wordpress authoring tool very quickly, as it's just not suited for serious scientific authoring. Nor does it provide good facilities for collaborative editing; normally, only one person can see a draft post, so you cannot pass this around between several authors.
>
> http://process.knowledgeblog.org/3

The KnowledgeBlog site encourages people to use their current authoring tools and treat the KnowledgeBlog WordPress platform as a publishing and review system.

Others are more positive about WordPress as an editor. Martin Fenner, for example is a tireless promoter of the practice. And the Digress.it help recommends using WordPress to create content from scratch, the opposite of the advice coming from KnowledgeBlogs:

We recommend using the WordPress editor directly for a number of reasons:

- Multiple authors can easily collaborate on a single document;
- A complete revision history of the document is maintained with the ability to roll-back to earlier versions;
- This method produces a web-ready document, native to WordPress, and avoids the two-stage process of 're-publishing' on your Digress.it site; and
- You can easily embed video and other objects.

And then there's Annotum. The site says:

Annotum will build upon the WordPress platform as a foundation, filling in the gaps by providing the following additional features:

- Rich, web-based authoring and editing:
  - "What you see is what you get" (WYSIWYG) authoring with rich toolset (equations, figures, tables, citations and references)
  - coauthoring, comments, version tracking, and revision comparisons
- Strict conformance to a subset of the NLM journal article publishing tag set

And a long list of other features. There is no code to show yet, though.

Collaboration platform

Others are seeing WordPress as a place for collaborative authoring and editing. Annotum promises this on a grand scale. For those who would like to get started, Martin Fenner listed some resources late last year:

The Co-Authors Plus Plugin enables multiple authors per article. Each author can be linked to an author page for displaying biographical info. WordPress could be extended to include additional info such as institution or past publications. Linking the WordPress user account to the unique author identifier ORCID, and describing the role of the author in the paper (e.g. conceived and designed the experiments or analyzed the data) would be particularly interesting. Plugins such as Edit Flow can extend the workflow by adding custom status messages (e.g. resubmission), reviewer comments, and email notifications.

http://blogs.plos.org/mfenner/2010/12/05/blogging-beyond-the-pdf/

Collaboration post publication is handled by a WordPress tool that's been a hit in the UK, and with JISC. Digress.it is a tool for public annotation and discussion of long-form documents. The JISC incarnation is at jiscpress.org. Digress.it is related to Commentpress. (They're different things although sometimes confused with each other at least by me. See them compared here.)

For a JiscPress example see this document, which has a number of comments.

Issues

Some issues I have observed with WordPress in the past include the problems with its authoring environment, covered above but also a number of other considerations.

There is the WordPress version of Microsoft's "DLL hell" - "Plugin hell" - many Wordpress plugins and/or themes interact with each other in unpredictable ways. I found this out first hand, trying to show-off some

work my team at USQ had done on an annotation system. It worked (with bugs) in a plain WordPress site, but failed completely in Martin Fenner's demo site where there are many other plugins installed. I never got to the bottom of that. Plugins also go out out sync with the WordPress as it evolves, so a site with lots of plugins can be hard to maintain, this is also the case with systems like Drupal which have their own enthusiastic following.

Some of the above systems require the content management system to be used in very particular ways – for example Digress it treats each document as a new WordPress site and asks you to upload posts in a particular order so that the Table of Contents for the site looks right. There are two issues with this kind of approach. I'm not saying that people are not already aware of these issues, but noting that they are there:

- There's sometimes a fair bit of overhead involved in setting things up just so. Sometimes, it would make sense to automate some of the processes. Other times maybe a re-think to reduce complexity might be in order.

- There is a risk of creating a new form of the proprietary lock-in we had up until recently (and arguably we still have) with document formats like Microsoft's .doc. The documents we create in some of these systems may end up being unusable in other systems. If you author a long document in Digress.it and depend on a particular configuration of WP and, having posts in a certain order and so on for the document's integrity, then it is essential to consider an exit strategy and an archiving strategy (more on that soon – an EPUB export might be just the ticket).

  There are similar issues/risks with stuff like WordPress shortcodes such as KCite from KnowledgeBlogs. It's a great tool for authors, allowing them to cite things in a rational way:

      DOI Example – [cite source='doi']10.1021/jf904082b[/cite]

      PMID example – [cite source='pubmed']17237047[/cite]

  But it's proprietary to a particular processing environment. If one wants to be able to re-used these documents or archive them then it is important to consider which version of the documents in WP to keep. (I'd argue that in this case best practice would be to transform the above to an RDFa representation in HTML and treat the HTML version as the version of record – more on this later in the project).

All this adds up to saying that **WordPress + plugins can be fragile** – the application itself needs to be updated frequently for security reasons, and so does the operating system underneath and inevitably stuff breaks. The more complex the plugin-set and the further you stray from straight WordPress the worse the risk. Even on simple sites there can be issues. For example one of the WordPress sites I use regularly currently has a bug with remote publishing via Atompub and XMLRPC. One day it was working and the next all my attempts to post from the tools I use everyday, as per the best practice advice from the KnowledgeBlog people, were minus the characters **<** and **>** in the document source, both of which are obviously essential to the web.

For those interested in learning more about WordPress for scholarship, there's a Google Group called WordPress for Scientists that is worth joining even if you are not a scientist and a test site that Martin Fenner has set up for WordPress plugins.