

How to add EPUB support to EPrints

[This is a repost from the JISCPub project – please comment over there
<http://jiscpub.blogs.edina.ac.uk/2011/05/03/how-to-add-epub-support-to-eprints-8/>]

In a [previous post here on the jiscPUB project](#) I said it would be good for the EPrints repository software to support EPUB uploads.

I'd love to do something with a repository – I'm thinking that it would be great to deposit theses in EPUB format – and the repository could provide a web-based reader, along the lines of [IbisReader](#), which Liza Daly and company created. I'm looking at you, Eprints! Eprints already almost supports this, if you upload a zip file it will stash all the parts for you in a single record. All we would need would be something like this [little reader my colleagues at USQ made](#). It would just be a matter of transforming the EPUB TOC into JSON, and loading the JavaScript into an Eprints page.

I Called Les Carr's attention to the post and he responded:

[lescarr @ptsefton](#) just tell us what to do and we'll do it.

OK. Here goes with my specification for how EPrints could add at least basic support for EPUB.

Putting EPUB into EPrints as-is

To explore this, I ran the [EPrints live CD](#) (livedcd_v3.1-x.iso) under VirtualBox on Windows 7 – this worked well when I gave it a decent amount of memory – it didn't manage to boot in several hours at 256Mb. (Note that no repositories were harmed in the making of this post – I did not change the Eprints code at all.)

The EPUB format is a zipfile containing some XHTML payload documents, a manifest, and a table of contents. On one level EPRINTS already supports this, in that there is support for uploading ZIP files. I tested this using Danny Kingsley's thesis (as received, with no massaging or adding metadata apart from tweaking the title in Word) [converted to EPUB](#) via the ICE service [I have been working on](#).

The procedure:

1. Generated an EPUB using ICE.
2. Changed the file extension to .zip.
3. Uploaded it into EPrints.

The result is an EPrints item with many parts. If you click on any of the HTML files that make up the thesis then they work as web pages – ie the table of contents (if you can find it amongst the many files) links to the other pages. But there is no navigation to tie it all together you have to keep hitting back – each HTML page from the EPUB is a stand alone fragment.

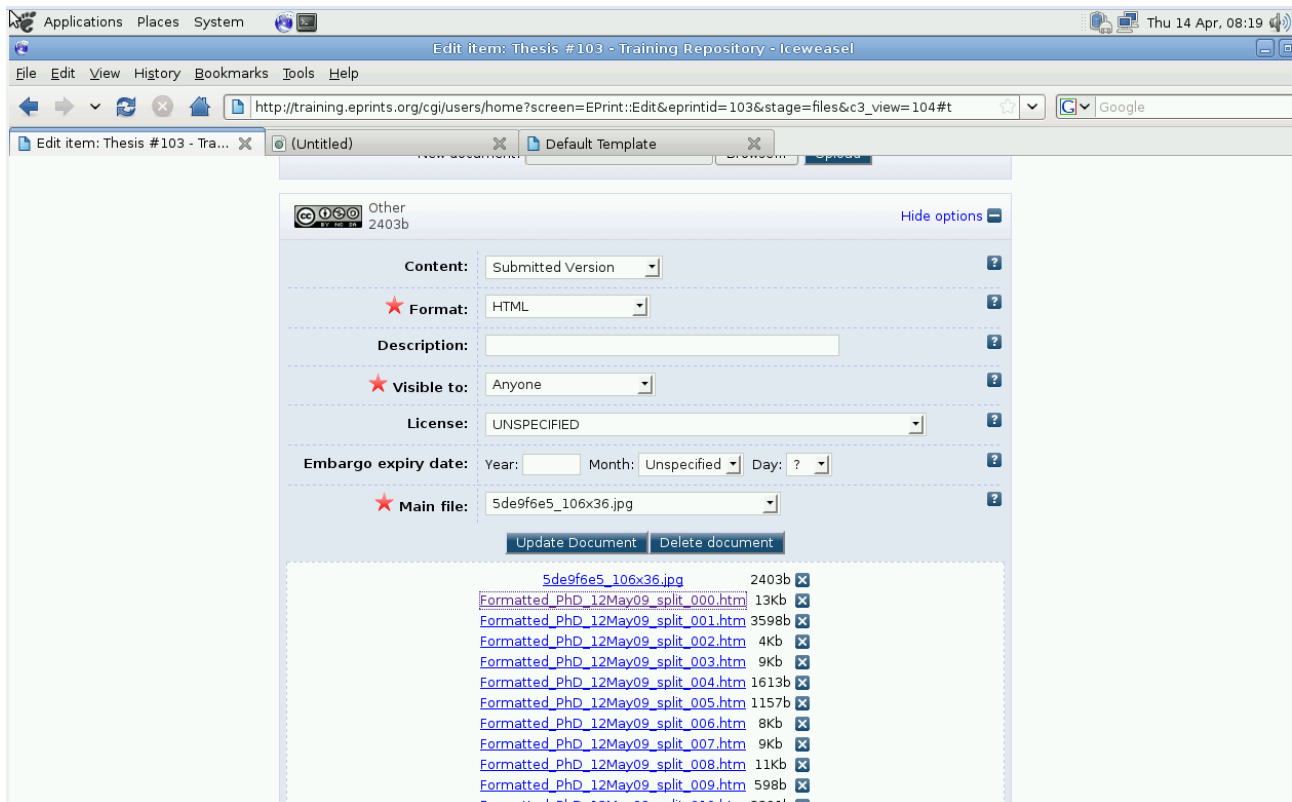


Illustration 1: The management interface in EPrints showing all the parts of an EPUB file which has been uploaded and saved as a series of parts in a single record.

At this point I went off on a side trip, and [wrote this little tool](#) – to add an HTML view to an EPUB file.

Putting enhanced EPUB into Eprints

Now, lets try that again with [the version](#) where I added an HTML index page to the EPUB using the new demo tool, epub2html. I uploaded the file, clicked around semi-randomly until I figured out how to see all the files listed from the zip, and selected index.html as the 'main' file. From memory I thought the repository would do that for me but it didn't. Anyway, I ended up with this:

The screenshot shows a web-based ebook reader interface. On the left, a vertical list of links serves as a table of contents, including 'Acknowledgements', 'Abstract', 'Glossary of Abbreviations', 'Table of Contents', and 'Chapter 1 - Introduction'. The 'Glossary of Abbreviations' link is selected, and its content is displayed in an inline frame on the right. This frame contains a table with two columns: the abbreviation and the full name of the organization or service.

AAP	American Association of Publishers
ACM	Association of Computing Machinery
ACS	American Chemical Society
ADFA	Australian Defence Force Academy
ANDS	Australian National Data Service
ANU	Australian National University
APAC	Australian Partnership for Advanced Computing

Illustration 3: A rudimentary ebook reader using an inline frame.

If I click on the link starting with Other, there we have it – more-or-less working navigation within the limits of this demo-quality software. All I had to do was change the extension from .epub to .zip and select the entry page, and I had a working, navigable document.

The initial version of epub2html used the unsupported epubjs as a web based reader-application – but Liza Daly suggested I use the more up to date Monocle.js library instead. I tried that but I'm afraid the amount of setup required is too much for the moment so what you see here is an HTML page with an inline frame for the content.

What does the repository need to do?

So what does the EPrints team need to do to support EPUB a bit better?

- Add EPUB to the list of recognised files.
- Upon recognising an EPUB...
 - Use a service like epub2html that can generate an HTML view of the EPUB. I wrote mine in Python, Eprints is written in Perl but I'm sure that can be sorted out via a re-write or a web service or something*.
 - Allow the user to download the whole EPUB, or choose to use an online viewer. Could be static HTML, frames (not nice), or some kind of JavaScript based viewer.
 - Embed some kind of viewer in the EPrints page itself, or at least provide a back-link in the document viewer to the EPrints page.

Does that make sense, Les?

[This is a repost from the JISCPub project – please comment over there

* Maybe there's a Python interpreter written in Perl?

<http://jiscpub.blogs.edina.ac.uk/2011/05/03/how-to-add-epub-support-to-eprints-8/>]

Copyright Peter Sefton, 2011-04-15. Licensed under [Creative Commons Attribution-Share Alike 2.5 Australia](http://creativecommons.org/licenses/by-sa/2.5/au/). <<http://creativecommons.org/licenses/by-sa/2.5/au/>>

This post was written in OpenOffice.org, using templates and tools provided by the [Integrated Content Environment](#) project.