

Anthologize: a WordPress based collection tool

[This is a copy of a post on the jiscPUB project – if you have comments please do so over there:
<http://jiscpub.blogs.edina.ac.uk/2011/05/11/anthologize-a-wordpress-based-collection-tool/>]

In this post I'll look at [Anthologize](#). Anthologize lets you write or import content into a WordPress instance, organise the 'parts' of your 'project' and publish to PDF or EPUB, HTML or into TEI XML format. This is what I referred to in my last post [about WordPress](#) as an aggregation platform.

Anthologize background and use-cases

Anthologize was created in an interesting way. It is the (unfinished as yet) outcome of a one-week workshop conducted at the Centre for History and New Media – the same group that brought us Zotero and Omeka, which is one good reason to take it seriously. They produce very high quality software.

Anthologize is a project of [One Week | One Tool](#) a project of the [Center for History and New Media, George Mason University](#). Funding provided by the National Endowment for the Humanities. © 2010, Center for History and New Media. For more information, contact infoATanthologizeDOTorg. Follow [@anthologize](#).

Anthologize is a WordPress plugin that adds import and organisation features to WordPress. You can author posts and pages as normal, or you can import anything with an RSS/Atom feed. The imported documents don't seem to be able to be published for others to view but you can edit them locally. This could be useful – but introduces a whole lot of management issues around provenance and version control. When you import a post from somewhere else the images stay on the other site, so you have a partial copy of the work with references back to a different site. I can see some potential problems with that if other sites go offline or change.

Let's remind ourselves about the use-cases in [workpackage 3](#):

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/>

Many documents like (a) theses or (c) reports are likely to be written as monolithic documents in the first place, so it would be a bit strange to write, say, a report in Word, or LaTeX or asciidoc (which is how I think Liza Daly will go about writing the landscape paper for this project), export that as a bunch of WordPress posts for dissemination, then reprocess back into an Anthologize project, and then to EPUB. There's much more to go wrong with that, and information to be lost than going straight from the source document to EPUB. It is conceivable that this would be a good tool for thesis by publication, where the publications were available as HTML that could be fed or pasted in to WordPress.

I do see some potential with (d) courseware here – it seems to me that it might make sense to author course materials in a blog-post like way covering topics one by one. I have put some feelers out for someone who might like to test publishing course materials, without spending too much of this project's time as this is not

one of the core use cases. If anyone wants to try this or can point me to some suitable open materials somewhere with categories and feeds I can use then I will give it a go.

There is also some potential with (c), project reports, particularly if anyone takes up the [JiscPress](#) way of doing things and creates their project outputs directly in WordPress+digress.it. It would also be ideal for compiling stuff that happens on the project blog as a supporting Appendix. So, an EPUB that gathers together, say all the blog posts I have made on WorkPackage 3 or the whole of the jiscPUB blog might make sense. These could be distributed to JISC and stakeholders as EPUB documents to read on the train, or deposited in a repository.

The retiring academic (b) (or any academic really) might want to make use of Anthologize too – particularly if they've been publishing online. If not they could paste their works into WordPress as posts, and deal with the HTML conversion issues inherent in that, or try to post from Word to WordPress. The test project I chose was to convert the blog posts I have done for jiscPUB into an EPUB book. That's use case (c) more or less.

How did the experiment go?

I have documented the basic process of creating an EPUB using Anthologize below, with lots of screenshots, but here is a summary of the outcomes.

Some things went really well.

- Using the control panel at my web host I was able set up a new WordPress website on my domain, add the Anthologize plugin and make my first EPUB in well under an hour. (But as usual, it takes a lot longer to back-track and investigate and try different options, and read the google group to see if bugs have been reported and so on).
- The application is easy to install and easy to use – with some issues I note below.
- Importing a feed just works if you search to find out how to do it on a standard WordPress host (although I think there might be issues trying to get large amounts of content if the source does not include everything in the feed).
- Creating parts and dragging in content is simple.
- Anthologize *looks* good.

The good looks and simple interface are deceptive, lots of functionality I was expecting to be there just wasn't – yet. I have been in contact with the developers and noted my biggest concerns, but here's a list of the major issues I see with the product at this stage of its development:

- There does not seem to be a way to publish the project (or the imported docs) directly to the web – rather than export it. Seems like an obvious win to add that. I can see that being really useful with Digress.it for one thing. The other big win there would be if the Table of Contents could have some semantics embedded in it so it could act like an ORE resource map - meaning that machines would be able to interpret the content. (I will come back to this idea soon with a demo of using Calibre to make an EPUB)
- There are no TOC entries for the posts within a 'part' that is, if you pull in a lot of WordPress posts, they don't get individual entries in the EPUB ToC.
- Links, even internal ones, like the table of contents links on my posts all point back to the original post – this makes packaging stuff up much less useful – you'd need to be online, and you lose the context of an intra-linked resource. [This is a known problem](#), and the developers say they are going to fix it.
- Potentially a problem is the way Anthologize EPUB export puts all the HTML content for the whole project into one HTML file – I gather from poking around with Calibre etc that many book readers need their content chunked into multiple files.

- There's a wizard for exporting your EPUB, and you can enter some metadata and choose some options – all of which is immediately forgotten by the application, so if you do it again, you have to re-enter all the information.
- Epubcheck complains about the test book I made:
 - It says the mimetype (a simple file that MUST be there in all EPUB) is wrong – looks OK to me.
 - It complains about the XHTML containing stuff from the TEI namespace and a few other things.
- Finally, PDF export fails on my blog with a timeout error – but that's not an issue for this investigation.

Summary

For the use case of bundling together a bunch of blog posts (or anything that has a feed) into a curated whole Anthologize is a promising application, but unless your needs are very simple it's probably not quite ready for production use. I spent a bit of time looking at it though, as it shows great promise and comes from a good stable.

Here's the result I got importing the first handful of posts from my work on this project.

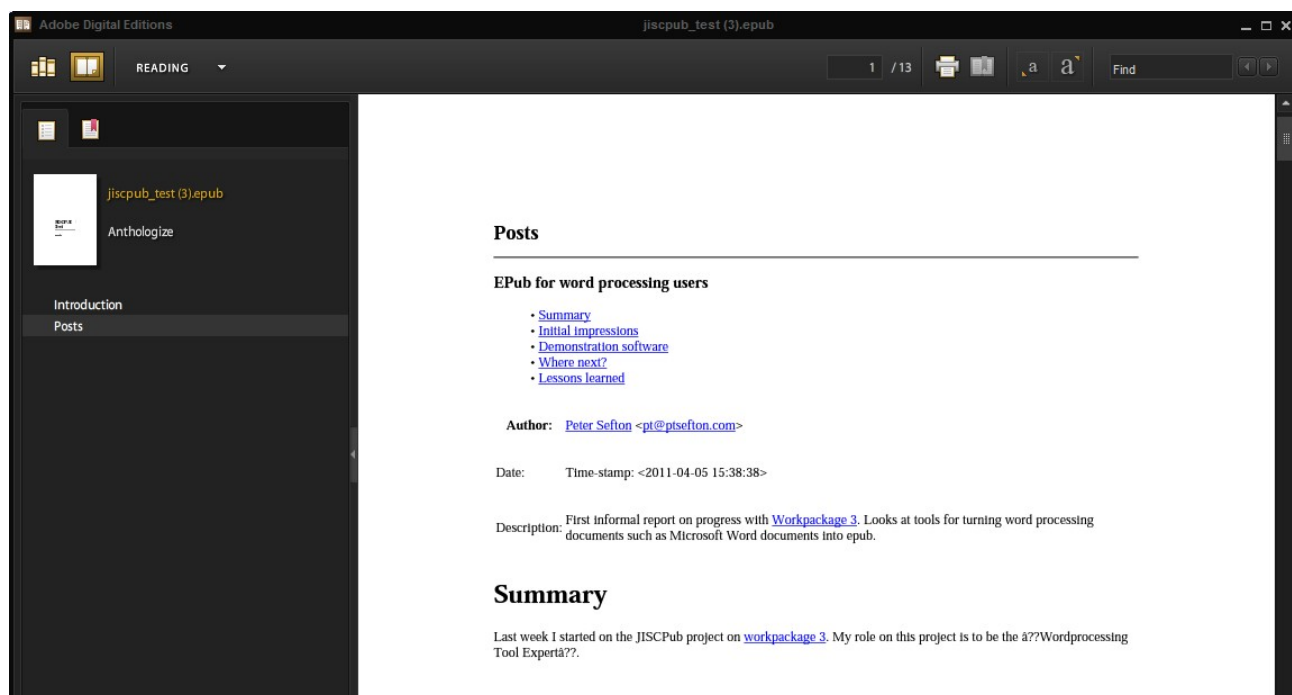


Illustration 1: The test book in Adobe Digital Editions - note some encoding problems bottom right and the lack of depth in the table of contents. There are several posts but no way to navigate to them. Also, clicking on those table of contents links takes you back to the jiscPUB blog not to the heading.

Walk through

Anthologize

My Projects [Add New](#)

You haven't created any projects yet.

[Start a new project.](#)

Illustration 2: Anthologize uses 'projects'. These are aggregated resources, in many cases they will be books but project seems like a nice media-neutral term.

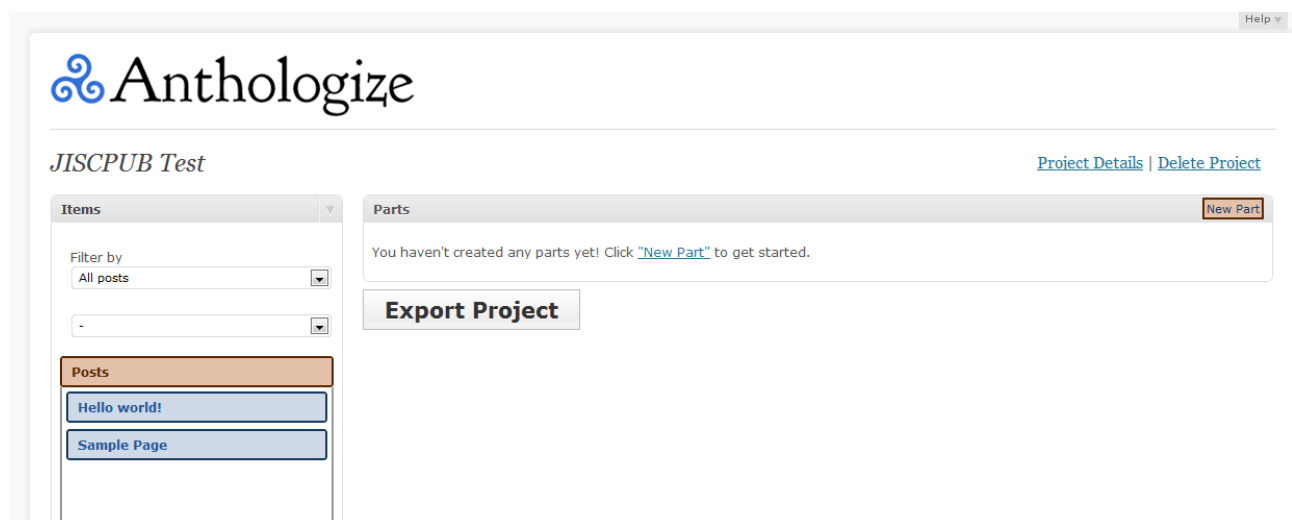


Illustration 3: A new project in a fresh WordPress install – only two things can be added to it until you write or import some content.



Import Content

Want to populate your Anthologize project with content from another web site? Enter the RSS feed address of the site from which you'd like to import and click Go.

Please respect the rights of copyright holders when using this import tool.

Feed URL:

Illustration 4: Importing the feed for workpackage 3 in the jiscPUB project.
<http://jiscpub.blogs.edina.ac.uk/category/workpackage-3/feed/atom/>



Import Content

Select the items you'd like to import to your Imported Items library and click Import.

Feed items:

- ☒ **How to add EPUB support to EPrints** 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 ... [Continue reading →](#)
- ☒ **Introducing Epub2Html – adding a plain HTML view to an EPUB** Background Demo Trying it out / the future Background EPUB ebook files are useful if you have an application to read them, but not everyone does. We have been discussing this in the Scholarly HTML movement; to some of us … [Continue reading](http://jiscpub.blogs.edina.ac.uk/?p=150) &arr;
- ☒ **Some questions about EPUB, WordPress, tools** If you publish EPUBs now, what tools do you use? What's considered best practice for EPUBs? JISC project people: What do you have to do to get your reports up in JISCPress? I have a couple of questions for discussion ... [Continue reading →](#)
- ☒ **Metadata in word processing monographs** Introduction – why worry about metadata? Thesis workflow Thesis metadata Summary Where now? [Reposted with spell checking a few minutes after original post, posted draft by mistake.] [And shortly after that added missing bibliography] Introduction – why worry about metadata? ... [Continue reading →](#)
- ☒ **Metadata in word processing monographs** Posted a draft version of this document by mistake. This post has been replaced by a new version <http://jiscpub.blogs.edina.ac.uk/2011/04/11/metadata-in-word-processing-monographs-2/>
- ☒ **EPub for word processing users** Summary Initial impressions Demonstration software Where next? Lessons learned Author: Peter Sefton <pt@ptsefton.com> Date: Time-stamp: <2011-04-05 15:38:38> Description: First informal report on progress with Workpackage 3. Looks at tools for turning word processing documents such as Microsoft Word documents into ... [Continue reading →](#)

Illustration 5: You can select which things to keep from the feed. Ordering is done later. Remember that imported documents are copies, so there is potential for confusion if you edit them in Anthologize.



Export Project

Select a project... JISCPUB Test ▼

Copyright Information

Year 2011

Copyright Holder Peter Sefton

Type ☐ Copyright ☒ Creative Commons Attribution ▼

Edition test

Add Author(s) Peter Sefton

Next

Illustration 6: Exporting content is via a wizard, easy to use but frustrating because it asks some of the same questions every time you export.

Export Project

Title

Dedication

This book is dedicated to David Flanders who organized the funding for this project via JISC>

Acknowledgements

I would like to acknowledge the family dog, Bootsy who kept my feet warm all afternoon.

Export Format

- ☐ PDF
- ☒ ePub
- ☐ HTML
- ☐ Anthologize TEI

Next

Illustration 7: Having to retype the export information is a real problem as you can only export one format at a time. Exported material is not stored in the WordPress site, either, it is downloaded, so there is no audit trail of versions.

[This is a copy of a post on the jiscPUB project – if you have comments please do so over there: <http://jiscpub.blogs.edina.ac.uk/2011/05/11/anthologize-a-wordpress-based-collection-tool/>]

Copyright Peter Sefton, 2011-05-04. 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.