Audiobooks in the hybrid publishing workfow

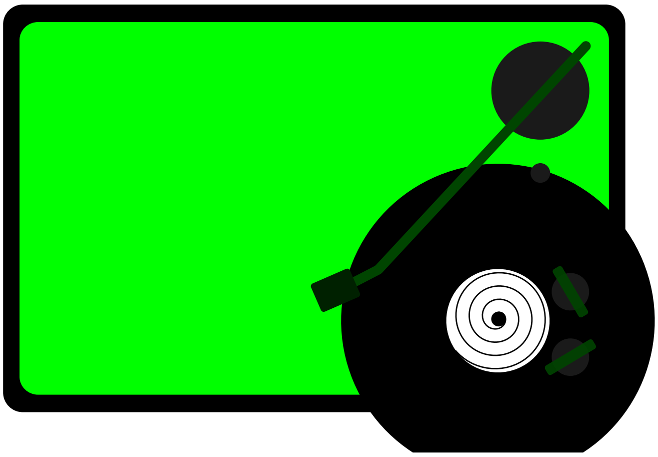
BLOG:

Audiobooks in the Hybrid Publishing Workflow

By Lucia Dossin (http://www.publishinglab.nl/blog/author/lucia-dossin/), January 30, 2017 at 9:32 am.

A bit if context

This is the last of a series of posts that belong to a research aimed at automated production of an audiobook within the Hybrid Publishing Workflow. Previous posts revolve around the first audiobooks (http://www.publishinglab.nl/blog/2016/04/13/mary-had-a-little-lamb/), text-to-speech software inquiries (http://www.publishinglab.nl/blog/2016/05/23/text-to-speech-software/) and a semi-automated approach to the task and its manual counterpart (http://www.publishinglab.nl/blog/2016/07/12/craft-your-own-audiobook/).



The Hybrid Publishing Workflow

The HPW (https://github.com/DigitalPublishingToolkit/Hybrid-Publication-Resources) relies on a makefile (https://github.com/DigitalPublishingToolkit/Hybrid-Publishing-Resources/blob/master/makefile) containing recipes to generate different outputs:

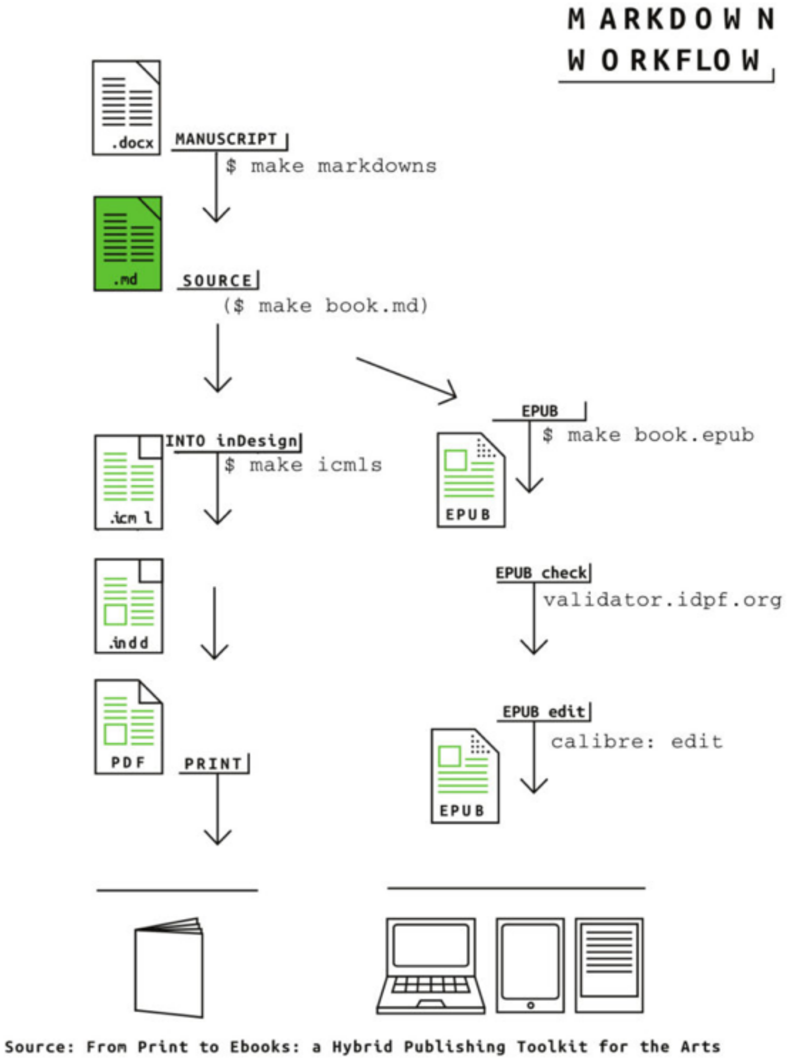
```
make markdowns
```

```
make epub
```

```
make icmls
```

and so on. For further details on this process, check From Manuscript to EPUB (https://github.com/DigitalPublishingToolkit/Hybrid-Publishing-Resources/wiki/From-Manuscript-to-EPUB) and Silvio Lorusso's workshop report (http://www.publishinglab.nl/blog/2017/01/10/make-book-make-soup-workshop-report/%20).

My goal was thus to produce a recipe for making audiobooks, which should fit the existing workflow structure, satisfy the specificities of an audiobook and its making process and be as user-friendly as possible.



The workflow

Audiobook and EPUB

My starting point was the EPUB making process in the workflow.

An EPUB is a collection of zipped text files – basically HTML, XML, CSS. It is therefore relatively fast to produce and easily editable. One can generate an EPUB and then edit details (from content to metadata, including cover) using an application like calibre. There is no strict need to re-generate an EPUB after edits thus — simply saving the edited file(s) is enough. Chapters can be identified according to markup, which means you can wrap all chapters in one text file — or keep them as separate files — and will still see them as chapters in the final output. Book structure is accomplished via markup.

Summing up, EPUBs are:

- fast to generate
- easy to edit (content and metadata)
- structured by markup

An audiobook does not share these properties. Converting from text to speech is a long-lasting process: during this research I used different computers and got very different results in terms of time needed to synthesize speech from text. But no matter how fast the computer is, the amount of time needed for this task is always much bigger than converting text from one format to another: many minutes versus a few seconds.

Audiobook metadata can be relatively easy to edit but editing content requires stepping back in the process and repeating text-to-speech conversion.

The chapter structure is achieved by a metadata file obtained by calculation (the algorithm must know the chapters files, its orders and durations and to be able to store the values in a file or variable). In short, structure needs to be calculated apart, it is not determined by markup. This also means that edits in content require new calculation of the chapters structure.

- not fast to generate
- content not easily editable (time consuming)
- structure by calculation

Also, while an EPUB requires pandoc, calibre and a code editor to be produced, an audiobook requires not only pandoc and a code editor as well, but also flite, sox, ffmpeg, MP4Box and mp4chaps. It is clear that these processes are very different in number of steps and complexity.

Reading Experiences

Not only technical aspects differ – the reading experience is also completely varied and with that some types of content can become more (or less) relevant within the publication. The translation of images and graphics into text represents a challenge not necessarily from a technical point of view (currently, image recognition software is relatively accurate and even a simple solution like the use of well-written alt tags will have the job done) but mainly from an experiential one. How accurate can the description of a color be? So, even though I consider this particularity to have been satisfactorily handled by the script, I see it as a topic that deserves much more attention and time — if one wants to seriously attempt to translate visuals into spoken words.

Some specificities belonging to the realm of the reading experience have not yet been satisfactorily tackled: for example, the speech synthesis mechanism used by the script is language-based. This means that foreign words and names will not be pronounced correctly. Also, listening to long URL's can be a bit disorientating. While the former can eventually be improved with technology, the latter is another good example of 'translation incompatibility', similar to the issue of describing a color with words.

Footnotes represent another remarkable difference. The EPUBs produced within the HPW will render them at the end of the text. But when one is listening to a text, it makes more sense to announce the reference immediately. The script encompasses this specificity by placing the notes inline and announcing them by prepending the word 'reference'.

Try it out: here you can download the EPUB version (http://networkcultures.org/wp-content/uploads/2015/12/TOD17-GrayZones.epub) and the audiobook of the publication The Gray Zones of Creativity and Capital (http://www.publishinglab.nl/wp-content/uploads/2017/01/TheGrayZonesOfCreativityAndCapital.m4b.zip).

The user

In order to minimize user frustration in the whole process – from installing necessary software to using the script, I provided a series of descriptive messages in the script and a tutorial (https://github.com/DigitalPublishingToolkit/Hybrid-Publishing-Resources/blob/master/README.md#audiobook) on how to install the software needed to produce audiobooks. These were improved according to feedback given by test users.

The script

The code is available in the Hybrid Publishing Resources repository (https://github.com/DigitalPublishingToolkit/Hybrid-Publishing-Resources).

By (opening the Terminal, navigating to folder containing the Makefile and) typing **make audiobook**, the makefile is executed and the audiobook recipe is run. Markdown files (placed inside the *md* folder) are required and should be named according to the order in which they should appear in the book. The metadata file (.yaml) is generated programmatically.

For the curious geeks out there: the recipe is essentially a python script that makes extensive use of the subprocess (https://docs.python.org/2/library/subprocess.html) module. Something that via the command line would be done like this:

```
pandoc -f markdown-inline_notes inputFile.md -t plain -o
outputFile.txt
```

Becomes:

```
pandoc_args = [ 'pandoc', '-f', 'markdown-inline_notes',
'inputFile.md', '-t', 'plain', '-o', 'outputFile.txt']
subprocess.check_call(pandoc_args)
```

If a call to subprocess fails, a descriptive error message will tell the user what to do. Like:

```
pandoc_args = [ 'pandoc', '-f', 'markdown-inline_notes',
'inputFile.md', '-t', 'plain', '-o', 'outputFile.txt']
try:
  subprocess.check_call(pandoc_args)
except:
  print 'Error. Please make sure that you have PANDOC installed
on your machine.'
  print 'Check http://pandoc.org/installing.html'
```

At some point close to the end of the process, the user is asked to input some metadata: Book title, Author title, Year and location of the file to be used as cover. The duration of the entire process can vary immensely, according to the length and number of text files and the processor speed.

Here is a demo of the process. Follow along!

For this demo I used a shortened version of the book chapters. The demo book duration is 44 minutes (made in around 2 minutes), while the original book is longer than 5 hours (made in around 12 minutes). Be aware that production time can vary immensely according to the computer's processor.

There is, of course, room for improvement and discussion. Comments and feedback (mailto:mail@luciadossin.net) are always appreciated.

SHARE THIS:

(http://www.publishinglab.nl/blog/2017/01/30/audiobooks-in-the-hybrid-publishing-workflow/?share=twitter&nb=1)

(http://www.publishinglab.nl/blog/2017/01/30/audiobooks-in-the-hybrid-publishing-workflow/?share=facebook&nb=1)

G+ (http://www.publishinglab.nl/blog/2017/01/30/audiobooks-in-the-hybrid-publishing-workflow/?share=google-plus-1&nb=1)

Tags: audiobook (http://www.publishinglab.nl/blog/tag/audiobook/), hybrid publishing (http://www.publishinglab.nl/blog/tag/hybrid-publishing/), text-to-speech (http://www.publishinglab.nl/blog/tag/text-to-speech/)

SHARE f (http://www.facebook.com/share.php?u=<full page url to share>) (https://twitter.com/share?text=Audiobooks in the Hybrid Publishing Workflow%20-%20) (mailto:?subject=Audiobooks in the Hybrid Publishing Workflow&body=http://www.publishinglab.nl/blog/2017/01/30/audiobooks-in-the-hybrid-publishing-workflow/.)

MOST RECENT READINGS