Creating Course Transcripts With Calibre

Preparing to Edit

Getting the Program

Calibre is supported widely, is free and open source, and can be downloaded here:
http://calibre-ebook.com/download

Note: If you'll be creating transcripts with LaTeX content, you must use a 3.X version of Calibre. Later versions will not work with the process described below. You can [download version 3.46 here](). It's possible to have multiple Calibre installations on one computer, so if you already have a later version, you can keep it and still install 3.X.

Once downloaded, it will automatically create a "Calibre Library" folder on your computer. This is where all your source files and exports will be held by default.
Download the two files "eC_Baseline-iframe-killa.txt" and
"eC_Baseline_Removal_Definitions_iframe-killa.csr" [here.]()

Getting the ePub

The epub export of the course content will probably be attached to the Wrike task for "Create Course Transcript;" if so, download it from there.

If it isn't in the Wrike task, ask the person who assigned you the task for the epub. Or, if you have the right Canvas permissions, follow these steps to download the epub from the course:
1. Log into your account on Canvas.
2. Navigate to the course instance you wish to generate your ePub from.
3. Click "People" and add yourself to the course as a Student.
4. Accept the invitation - you'll see a link you can click, or go to the course homepage and accept.
5. Click Account > Settings > Download Course Content
6. Find the course you want to generate and click "Generate ePub". When it finishes generating, click "Download ePub." Some courses may also show a "Download Associated Files" folder. Download this too and see what the files are. Typically, it's extra files that aren't used in the course. If so, you can ignore them. In almost every case, you won't need to do anything with these files to create the transcript.
7. The ePub file will automatically save to your Downloads folder or wherever your browser is configured to save downloads.

Loading the ePub into Calibre
1. Open Calibre from your Applications folder
2. To add the ePub to your Calibre library, just drag it into the Calibre window and it will now be mirrored in your Calibre library

Setting up Calibre Preferences

It's handy to have the File Viewer and File Preview enabled in the Calibre editor. Viewer shows you a list of the .xhtml files that make up the epub. Preview lets you see what the page will look like, approximately, when you export it to PDF.

If they're not showing, enable them by going to Tools in the top toolbar and checking File Preview and File Viewer.

Be sure to save your work often by clicking the floppy disk icon in the top toolbar; otherwise your changes will be lost.
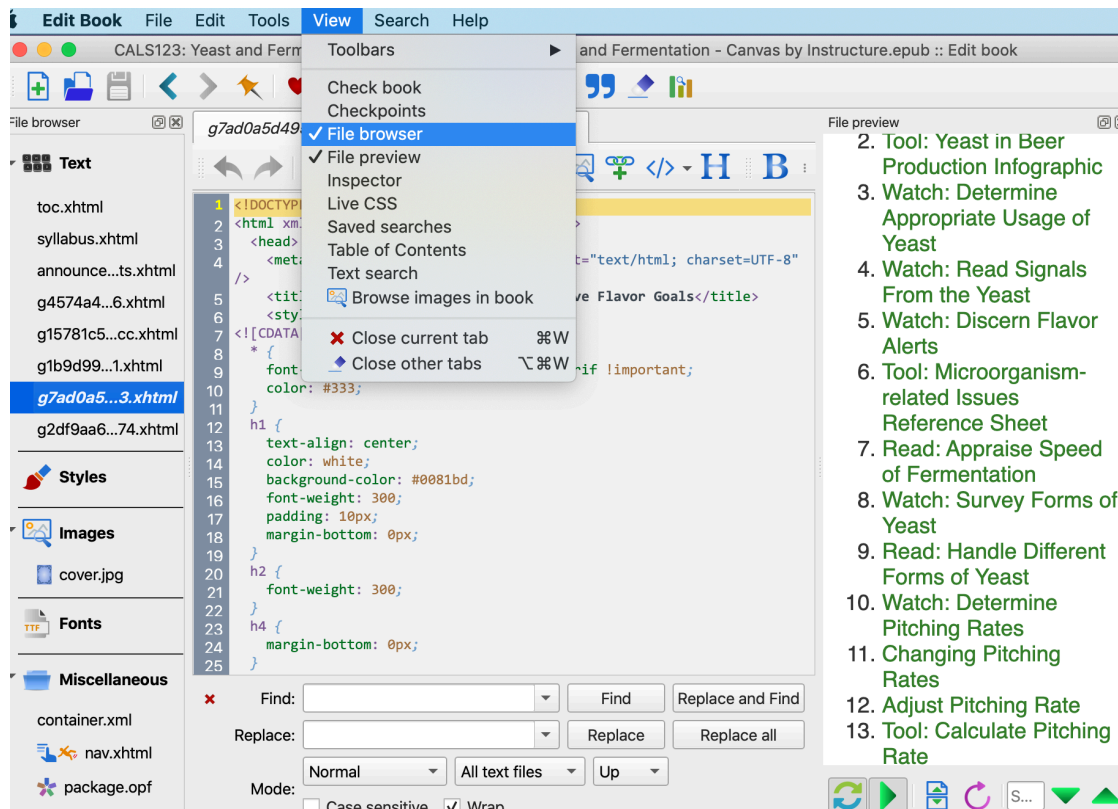
Editing in Calibre

Removing the Extraneous Content

1. Right click on the ePub you just added to Calibre and select "Edit Book"
2. There are usually several .xhtml child files that you should delete: "**syllabus.xhtml**", "**announcements.xhtml**", "**student-lounge.xhtml**", and the **course resources file** (this might be called "course-resources.xhtml or it might have a random name - open it and check the title). Remove them by right-clicking them in the File Viewer and choosing "Delete selected file".

3. Delete the entries for the above pages in any Tables of Content they appear in. Each page will be a list item (in HTML, **
** THE ITEM **
**) in the main Table of Content in toc.xhtml and will also be a list item in the ToC for the module it's in.

For example:
- *Click toc.xhtml in the Calibre editor*
- *Scroll down the page until you see this:*

   `<li>`

`<a href="#stay-connected">Stay Connected</a>`

`</li>`

- *Delete the code highlighted above.*
- *Repeat this process in the table of contents in the .xhtml file that the Stay Connected page appears in, probably the last module in the course.*

4. In the content of the.xhtml files, delete these pages, if they're present: Meet Your Class, Student Lounge, Stay Connected, Course Exit Survey.

   Individual pages are delineated by the **page title inside an <h2> tag at the top**, and "**<div style="page-break-before:always;"></div>**" at the bottom.
   For example, here is all the code for the Stay Connected page:

```
<h2 id="stay-connected">Stay Connected</h2>
      <p></p><h2>Stay Connected</h2>
      <p><iframe src="https://www.ecornell.com/canvas/" width="100%"
      height="1000"></iframe></p>
          <hr /><p><a href="#g7ad0a5d495addc6f2d2545f6bcead3c3.xhtml">
           Back to Module 3: Handle Yeast to Achieve Flavor Goals
          </a></p>
          <p><a href="toc.xhtml">Back to Table of Contents</a></p>
      <div style="page-break-before:always;"></div>
```

   To delete a page from an .xhtml file, highlight all the code from the start of the page - the <h2> tag down through and including the page-break-before div, then delete it, and save your work.

Adding the Video Transcript Text

You'll need to add the text transcript for each video to their corresponding pages in the xhtml files in Calibre.

The video transcripts of the course you're working on appear in the "transcript widget" — the pane that appears directly under the video player that shows the transcript text. You will copy the transcript text from the widget and paste it into the Calibre file.
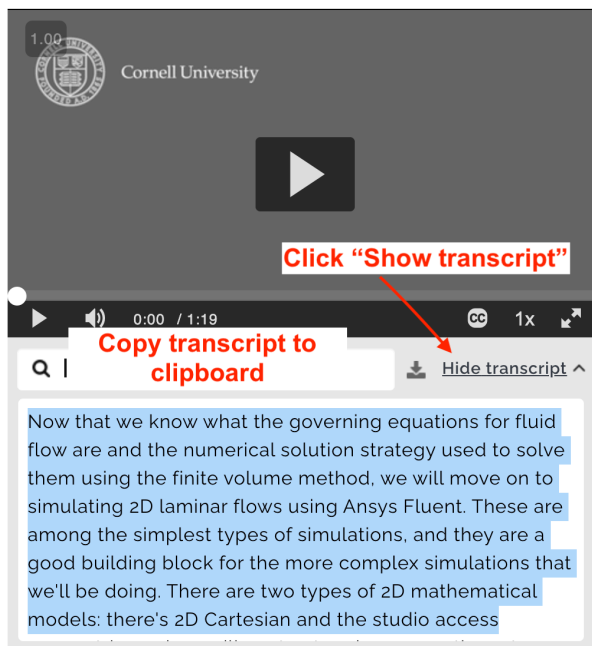
If a video doesn't have the transcript widget, make a note of it and move on. The person who assigned you the transcript task will help find the missing content.

Process to Add Video Transcripts
1. In the Calibre "Edit" dialog, open the first .xhtml file that correlates to a course module — the filename will be a random character string, like gbf13200e867mm09frb.xhmtl.
2. One way to find the videos is to search for class="entryId" . Each time that appears, there's a video that needs a transcript. You'll see a <div> tag wrapping a tag. You'll add the video transcript after the closing </div> tag.
3. Copy this code snippet (it's handy to keep it on a text file or other place where it's easy to copy and paste it when needed):
   <h3>Video Transcript</h3>
   <p></p>
4. Paste the snippet after the closing </div> tags that define the video player.

*Example***:**
<div id="kaltura1" class="kaltura_video resp2" style="margin: auto; width: 640px; height: 390px; margin-bottom: 1em; background: #ccc;">1_ug152clm</div>
<h3>Video Transcript</h3>
<p>Here's my lecture. Today I will be talking about Spanish wines.</p>

5. Now you'll locate the video transcript text for that video. Go to the course in Canvas, and find the corresponding page. Click Show Transcript on the video player, and copy the transcript from the widget.
6. Paste in the transcript text inside the <p></p> tags. If the transcript is longer than 3 minutes or so, add some paragraph tags to break up the text, for readability on the page.
7. Do this for all the Watch pages, Ask the Expert pages, and any page that has a video in all of the module .xhtml files
8. **Pages with multiple videos**: You'll probably see a lot of <div>s and extra code on multiple-video pages. You'll need to delete the extra code because it doesn't render correctly in the final PDF. Instead, just line up the videos one after the other. If there's content like a question or an expert's name, you can wrap them in <h3> tags and they should display nicely.
    ○ On ATE pages, format the questions as <h3>QUESTION</h3>
            For example, <h3>How can people begin to make change?</h3>
            <p>People can make change by…..</p>
    ○ To organize the video transcripts, for ATEs put <h3>EXPERTNAME</h3> if there are multiple experts, and list their answers under their name.
    ○ For tabbed content, title the video transcript with <h3>TABTITLE</h3>, then list the tab's answers below the <h3>. The tab titles appear in the html in the fragment tags.

Handling HTML Exceptions

It's beyond the scope of this guide to describe how to handle every possible HTML configuration you might find in a course. You might see Show/Hide toggles, multiple content tabs, large complex tables, and so forth. These things might look fine on the course page, but won't render correctly — or might not make sense — in the PDF format.

When you need to format this kind of code, remember that the aim is to provide the course content to the student in a clear, easy-to-read way, not to create an absolute mirror image of the course pages. You may need to use your own judgment and skills when working with some pages.

For example, a page may say something like

> *Think about Question X on your own. When you think you have the answer, click the icon below to read Professor Y's answer. <ICON>*
> *<show/hide toggle code>Answer.</show/hide close tag>*

Clicking an icon to hide/show the answer works on a webpage, but doesn't make sense in a PDF. So you might rework it something like this:

> *Think about Question X on your own. When you think you have the answer, read on for Professor Y's answer.*
> *<YOU REMOVE THE HTML ICON CODE FROM THE EPUB> Answer.*

When you come across pages or chunks of code that don't render nicely in a file, do your best to rework them so they read in a logical, clear manner. If you get stuck, feel free to ask Karen or another member of the CT team — Markette Pierce or Maeve O'Hara — for help.

Adding the Homepage and Configuring ToC links

1. You'll add the homepage content to the top of the "toc.xhtml" file. This will "break" the Table of Content links at the bottom of each page, so you'll need to fix that by doing the following:
2. Go to the first module file and pull up "Find and replace." Enter the following:
    a. Find: <a href="toc.xhtml">
    b. Replace: <a href="toc.xhtml##main_nav">
    c. Click Replace All
    d. Do this for the other module files

3. Still in the toc.xhtml file, change ordered lists in the Table of Contents to unordered lists. To do this, pull up Find and Replace, and:
    a. Find
    b. Replace
    c. Find
    d. Replace
    e. Do this for each module file as well, *but in the module files, verify that you're only changing the table of contents s, not any lists that appear in the course content.*
4. At the top of the toc.xhtml code just underneath the <body> tag, add this: <h1>(coursetitlehere)</h1>
    a. *Don't include the course code:*
    ● Correct: <h1>Generating User-Centered Solutions</h1>
    ● Incorrect: <h1>CESYS514: Generating User-Centered Solutions</h1>

5. Now you'll add the homepage content to the top of the "toc.xhtml" file. Go to the course homepage in Canvas and click Edit.
6. Go to HTML edit mode. Copy the code starting at the "What you'll do" section, and down to the end of the faculty bio.
7. Paste this code above the first <h1> tag that appears after the <body> tag in the "toc.xhtml"

8. Remove all the div tags in the code that you pasted - so delete all <div blah blah> and </div blah blah> tags, just in the chunk of code that you just pasted in.
   a. Be sure not to delete the faculty author image that usually appears right after the Faculty overview div tag. Delete the divs that surround the image tag but leave the image there.
9. Add this right before the Table of Contents tag:
   a. <div style="page-break-before:always;"></div>
10. Still in the toc.xhtml file, add the homepage video transcript at the bottom of the homepage, under the faculty bio. To get the video transcript, copy the transcript from the "widget" underneath the video player.
11. Once you have the transcript copied to your clipboard, add the below code and the transcript. Like so:
    (end of faculty author's bio) ...of the undergraduate curriculum.</p>
    <h3>Welcome Transcript</h3>
    <p>Homepage transcript goes here.</p>

12. **Do not include the Start Course link**. It will point to the online course, which students will not have access to after the course ends. Delete the Start Course code from the homepage, if you pasted it into the toc.xhtml file.

13. Save the file and close the editor

Downloading Course Images

Note: This section applies to Calibre versions 4.X and later. If you are using version 3.X or earlier, you can skip this section.
In order for course images to show up in the PDF transcript, you need to download them and save them to Calibre:
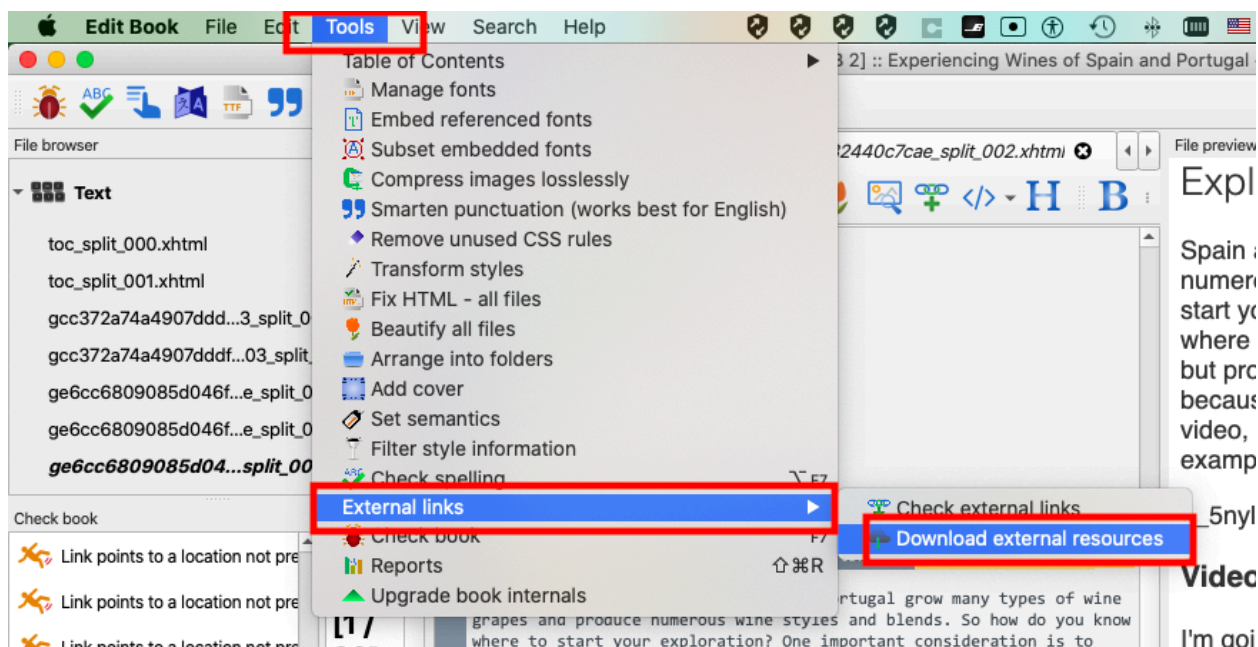
- In the Editing interface, at the top of the screen, click Tools -> External links -> Download external resources.
- Click Save.
- You should see a confirmation message sa
- To verify whether the images have downloaded successfully, select the last module file in the editor - whichever module has the Thank You and Farewell page. (You're selecting the TY page only because there will be an image on it - the faculty author photo. You can use any other page that you know has an image if you want.)
- Click File Preview. In the preview that appears, make sure the faculty photo shows up. If it doesn't, try these steps again.

Checking for Errors

There are two error-check steps you must take before converting epub to pdf. Best to check for errors before converting, as you can convert a whole file only to find it didn't format correctly due to inline errors.

Step 1 In the Calibre editor, click on the bug icon and you will get a list of possible errors within the epub, along with a general and specific link to the spot in the html where the error is occurring. *You just need to fix HTML parsing errors.* You can ignore errors with "duplicate" in their description, and "Link points to location not in file."

Step 2. Double click on the ePub to open the eBook viewer. Skim over the document and look for red boxes that say "This page contains the following errors". If these are present, you need to reopen the editor and fix them. The errors will specify what needs to be fixed.

Fixing LaTeX Equations

Important note: If you're creating a transcript with LaTeX content, you must use a 3.X version of Calibre. Later versions will not work with the process described below. You can [download version 3.46 here](#). It's possible to have multiple Calibre installations on one computer, so if you have a later version, you can keep it and still install 3.X.

In courses that contain LaTeX (mathematical) equations in the course text in Canvas, these additional steps need to be taken to get the equations to appear in the PDF. Without following these steps, the equations will appear as utter gobbledygook.

1. Open the ebook editor for the epub that you want to fix equations for
2. Open any of the .xhtml documents in the ebook from the file browser on the left of the ebook editor
3. Press Command + F to open the find and replace feature
4. In the "Mode" field of Find and Replace, click the dropdown and select "Regex function"
5. Next to the "Mode" field, click the file selector dropdown and select the option "All text files"
6. In the "Find" field enter this **exactly:** `<img class="equation_image"(.*?)/>`
7. If you've never done this before, click the "Create/edit" button
   a. If you have done this before you can simply click the Function field dropdown and select "fixLatex"
8. In the resulting popup for creating/editing a function, set the function name as "fixLatex" and paste the contents of this script into the Code area: https://ecornell.s3.amazonaws.com/casey/calibre/calibreLatexFix.py
9. Click OK
10. Now, in the Find and Replace feature click "Replace all"
11. This should fix the vast majority of the equations so they render in the converted PDF
12. *After you convert the PDF, you should check all the pages for clearly broken equations.* If you find broken equations, see if there's a clear error in the HTML and fix it and re-convert. If you can't figure out how to fix it, submit one of [these requests forms](#) detailing the issue and attaching the relevant epub and we will try to resolve it ASAP.

Converting & Publishing

Configuring Your Conversion Settings

1. If you haven't already done so, download the 📄 eC_Baseline_iframe-killa.txt and eC_Baseline_Removal-Definitions_iframe-killa.csr files from the shared folder "Calibre Printables - Files & Instructions" and save them somewhere you will remember
2. Right click on your ePub in Calibre and select Convert Books > Convert Individually.
3. The settings you will edit are:
    a. Metadata
    b. Look & Feel
    c. Page Setup
    d. Search & Replace
    e. PDF Output
4. You don't need to touch the other settings.
5. **Metadata**
    a. Set the output format to .pdf in most cases. Your Wrike ticket will specify if you need to export to a different format such as .docx.
    b. Edit the metadata on the right in the Metadata fields:
        i. Authors: put the author's name in Last, First format. The author's name can be found on course homepage.
        ii. Publisher: Cornell University for non-LAW and non-LEGAL courses
        iii. Publisher: Cornell Law School for LAW and LEGAL courses
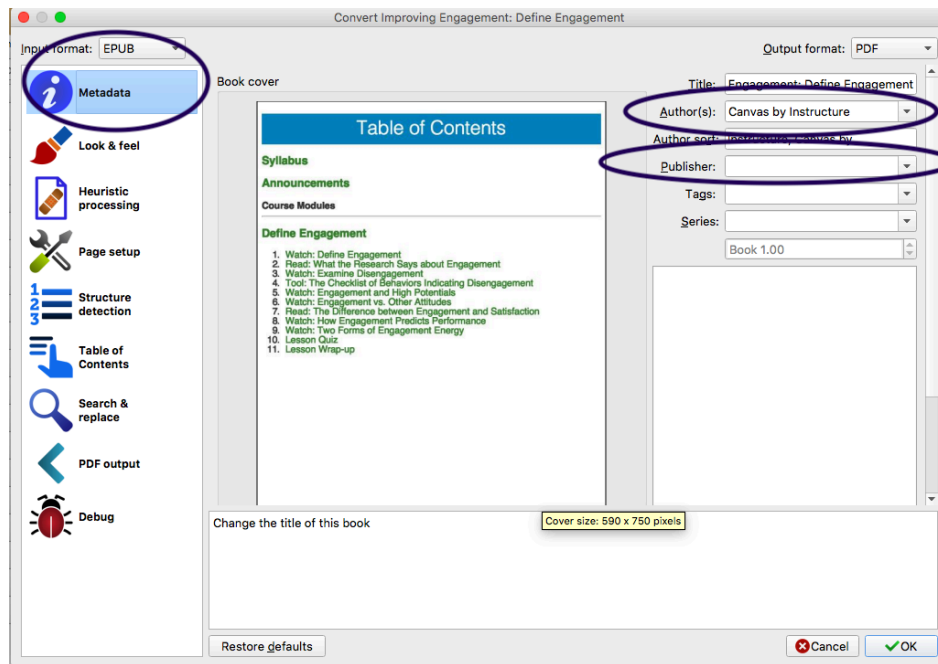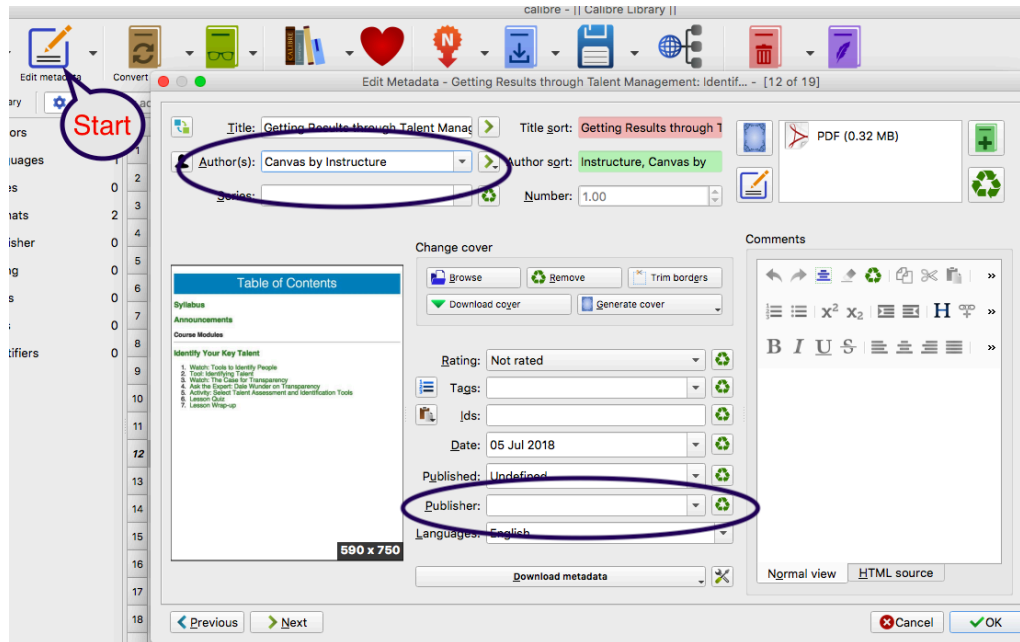
6. Look & Feel

   a. Select "Expand CSS"

   b. Go to "Transform styles"

   c. Select "Import"

   d. Import the file "eC_Baseline_iframe-killa.txt"

7. Page Setup

   a. Select "Default" for both the input and output profiles

   b. Set the margins to

      i. Left/right - 30.0 pt

      ii. Top/bottom - 50.0 pt

8. Search & Replace

   a. Click "Load"

   b. Select the file "eC_Baseline_Removal-Definitions_iframe-killa.csr"

9. PDF Output

   a. If using Calibre 4.X and earlier… add header and footer information in Calibre:

      i. **Header template- replace COURSE TITLE and SPONSORING UNIT with the correct values:**

         1. <p style="float:right; color: #3b3b3b; font-size: 53%;">COURSE TITLE
SPONSORING UNIT</p><p style="padding-top: 10px;"><hr style="border: 0; height: 0; border-top: 1px solid rgba(0, 0, 0, 0.1); border-bottom: 1px solid rgba(255, 255, 255, 0.3"/></p>

      ii. **Footer template:**

         1. <hr style="border: 0; height: 0; border-top: 1px solid rgba(0, 0, 0, 0.1); border-bottom: 1px solid rgba(255, 255, 255, 0.3"/><div

style="font-size: 46%; text-align:left; color: #3b3b3b;">© 2022 Cornell University.</div>

If using Calibre 5.99.8 or later…

iii. You'll need to check the box to "Use page margins from the document being converted" in order to maintain correct margins in the PDF.

iv. Leave the rest of the information on this tab alone. See the [Adding Header and Footer in Adobe (Calibre 6 beta)](#) section below.

10. Click OK to convert your file to pdf or other output format.

If you do not have access to Adobe Acrobat, S3, or access to the course as more than a student, at this point your job is done. Send epub and pdf to whoever gave you the assignment.

Adding Codio Exercise PDFs

Some technical and computer-based courses contain coding exercises in a programming interface called Codio. In order to include the Codio content in the transcript, you will need to download the content from Codio in PDF format, then insert them into the transcript file in the appropriate places. Here's the process for getting the Codio PDFs: https://docs.google.com/document/d/1QJprFUhpcOrh5VnNwg8kBWehGM8boXYa7Z8ZwqiI-8w/edit?usp=sharing

Final Refinement in Acrobat or Word and Upload to Course

1. Use Acrobat to remov
2. e any blank pages & the first autogenerated table of contents page (which will look fuzzy compared to the other pages and will have a blue header.)
3. Click through the course page by page and quickly review each for
   a. Keep an eye out for large blank spaces on non-watch pages. These may be missing images: double check against course or look in epub html in that spot, and add image if necessary.
   b. Inspect any tables that appear in the course - sometimes they run off the page. To fix this, first adjust any hard-coded column widths and recheck; if that doesn't work you'll need to get creative. Ask for help if needed.
4. Do not add the eCornell or Cornell seal or logo.

Adding the Glossary into the PDF

5. Navigate to the course folder in S3, or to the master course, and download the PDF glossary, if the course has a glossary. It will probably be toward the end of the course.
6. In Adobe, select "More Tools" and then "Combine Files"

a. Click "Add Files" and select the course transcript file and the glossary file. Then click the "Add Files" button to see them in the Combine Files view.
    i. Drag the files into the correct order, with the transcript on the left.
b. Click Combine
c. The new combined file will appear in its own tab with a title of Binder<number here>. Save it now as COURSECODE_course-transcript.pdf.
d. Using Acrobat inTools -> Content -> Link, select Link, select the text or area that should be the link, then in the popup that appears, select Page View or Open a Webpage, and then either paste in the URL, or scroll in the PDF to the page to jump to.
e. See the GIF for a demo: https://recordit.co/xOpkOVIIGt

Accessibility Remediation

In Adobe Acrobat or Word, run the built-in accessibility checker.
Click on any errors that show up, and if Fix is available as an option, do so.
Usually you can fix these errors: Tagging, Title, Language.
Sometimes you can fix heading level errors.
You usually cannot fix Alt Text in Figures or Elements errors.

Adding Header and Footer in Adobe (Calibre 6 beta)

When using Calibre 5.99.8 or later, you'll need to add the header and footer in Adobe rather than in Calibre.

1. From the Edit tab, click Header and Footer and select Add.
2. Use the following configuration:
    a. Font: Arial, size 8
    b. Margin (Inches): Top/Bottom: 0.5, Left/Right: 1.0
    c. Right Header Text:
        i. Course Title
           Sponsoring School
    d. Left Footer Text:
        i. © 2022 Cornell University.
3. Fill in the details above with course specifics and click OK.

Adding PDF to S3 and to Course

1. If you have access to S3, upload the edited epub and pdf to S3 (don't make them public).
2. Upload the epub and pdf to the Wrike task.
3. Upload them both to the course under "Files."

4. Add the transcript to "Course Resources section" as a File. Click edit and change name in Canvas to Course Transcript.
5. Click into the course transcript in course resources, right click on download option in upper left, and copy link address
6. On the homepage in the "this course includes" bulleted section:
   a. Add "One course transcript" as the last list item
   b. The homepage bullet text should read "One <link>course transcript</link>"
   c. The li class for the transcript should be <li class="transcript">
   d. Link the the transcript in the Files section of the course
   e. In the href tag, delete this code: class="instructure_file_link instructure_scribd_file"
   f. In the href tag, change the title attribute to title ="Course Transcript"
7. For On-Demand only, reference ILRHR552OD1. *You will probably not be working with an On-Demand course so can ignore this step.*

Handy Checklist for Process

Now that you've read through the process in detail and tried it out, here is a checklist you can copy and reuse if it helps your workflow, and as an easy way to check if anything major has been added/deleted from the process steps.

Keeping Rule Files Current

As we make more course transcripts, we may notice patterns in the Canvas html that should be stripped if present (for example, ATE divs). When these are noticed, it's best to update the .csr or .txt rule files and re-upload.

Tips and Tricks

Occasionally, one or more pages will be missing the page content, or a page might be missing completely from the epub. Usually this happens for pages with a special character in the title, like Read: Tips & Tricks, or Watch: Use "Agile" Methodology. Special characters are non-alpha-numeric characters, like ", &, ', :, and so forth.

To fix this, in the Canvas course, go to the appropriate page, enter Edit mode, enter HTML mode, and copy the html code for the missing content. Click 'Cancel." Then paste the content into the .xhtml file in the appropriate place.

END OF TUTORIAL!
Experimental - ignore below this point!

To add headers/footers in Calibre not PDF - for accessibility: