The Care and Keeping of eBooks

Rebecca Welzenbach, Digital Publishing Project Manager

rwelzenb@umich.edu

Find me on Twitter: @rwelzenb

Logistics

• We're in the Hatcher Library Gallery Lab

• Restrooms down the ramp and into the hall—women's on this floor, men's downstairs

• This workshop goes until 5 p.m. For the first half (or so), we'll mostly be learning about EPUB by looking at examples. In the second half, I hope you'll experiment—with my files, or your own

• Materials are available at
http://www.lib.umich.edu/digital-publishing-production/workshops

• We'll leave about 5-10 minutes at the end to fill out an evaluation. Access the evaluation at http://tinyurl.com/myES2011

Goals for today:

• Understanding of EPUB ebook format, how it fits into the landscape of digital publishing, and how the format works

• Awareness of EPUB conversion tools and how to use them as a starting point for generating EPUB

• Confidence to inspect, edit, and create EPUB of your own

• Awareness of self publishing channels that provide support for the creation and distribution of ebooks

Important caveat!

This workshop is meant help you do cool things with:

- Your own work to which you retain the rights
- Works in the public domain
- Works that you are otherwise licensed to copy, edit, and/or distribute.

Questions about copyright? Contact the library's Copyright Office: copyright@umich.edu
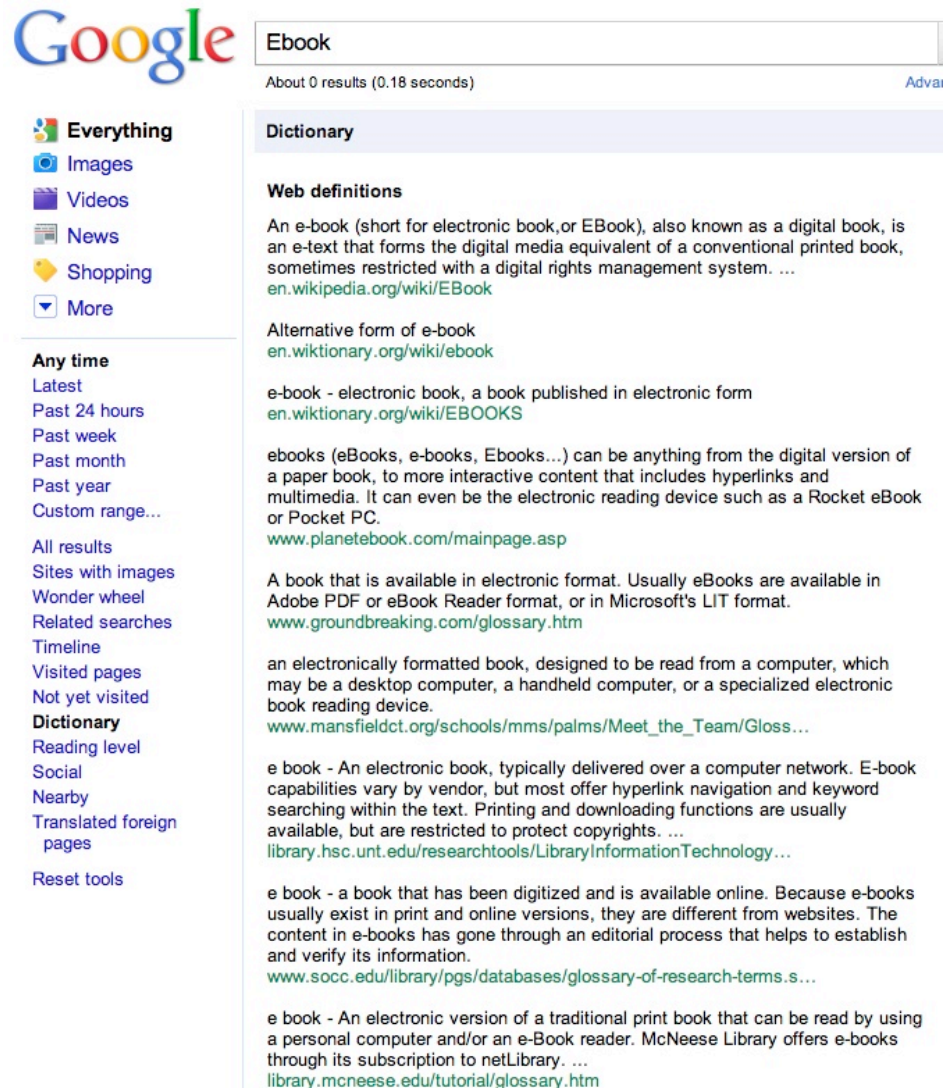
First things first….

Do those who are interested have….

• An EPUB reader? (try: http://ibisreader.com/)

• Text Editor? (Try Text Edit or Notepad, one of which is probably bundled in with your OS)

…or, look on with a neighbor and work together
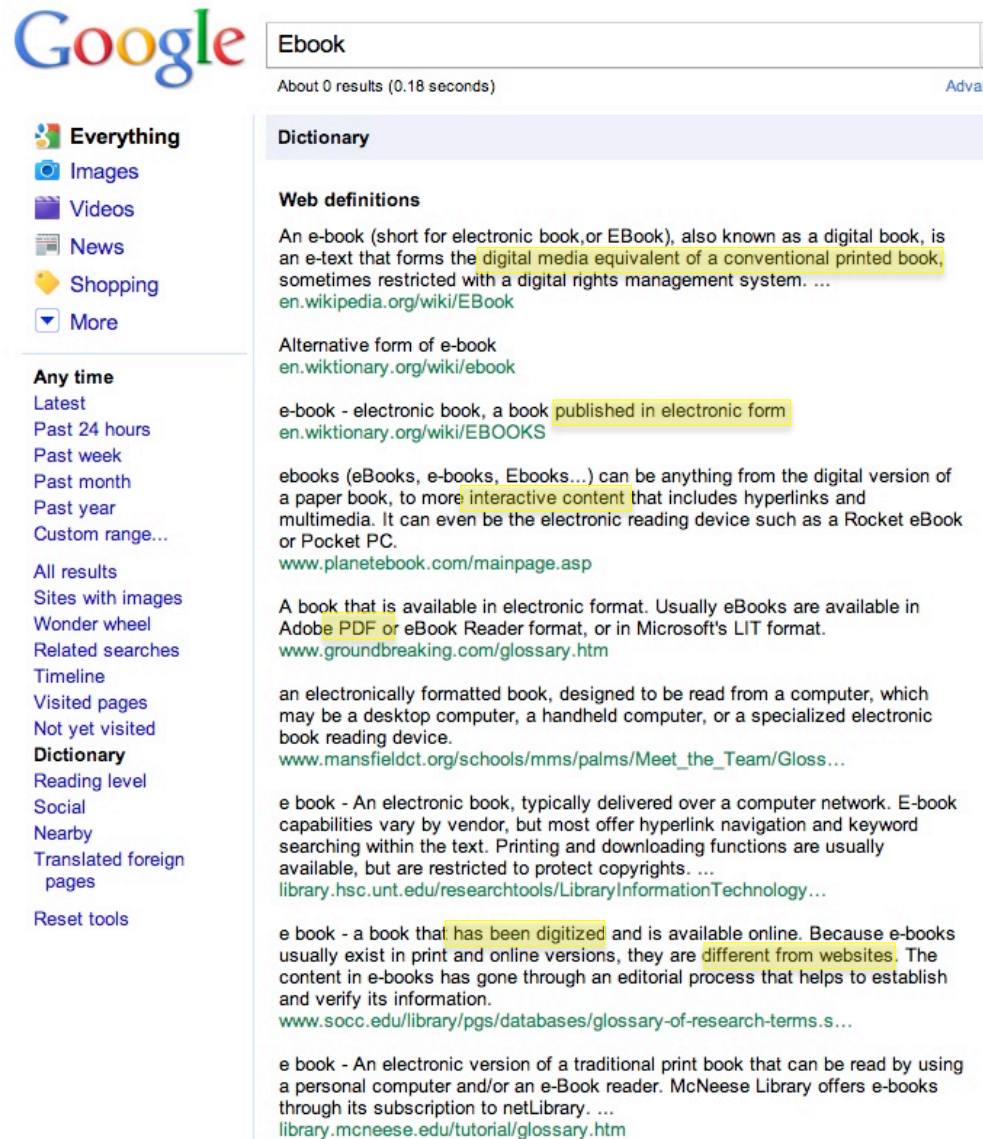
What is an ebook?

What is an ebook? A book you read electronically.

What is an ebook? A book you read electronically.

Google

Ebook

About 0 results (0.18 seconds)   Advar

- Everything
- Images
- Videos
- News
- Shopping
- More

Any time
Latest
Past 24 hours
Past week
Past month
Past year
Custom range...

All results
Sites with images
Wonder wheel
Related searches
Timeline
Visited pages
Not yet visited
Dictionary
Reading level
Social
Nearby
Translated foreign pages

Reset tools

Dictionary

Web definitions

An e-book (short for electronic book, or EBook), also known as a digital book, is an e-text that forms the digital media equivalent of a conventional printed book, sometimes restricted with a digital rights management system. ...
en.wikipedia.org/wiki/EBook

Alternative form of e-book
en.wiktionary.org/wiki/ebook

e-book - electronic book, a book published in electronic form
en.wiktionary.org/wiki/EBOOKS

ebooks (eBooks, e-books, Ebooks...) can be anything from the digital version of a paper book, to more interactive content that includes hyperlinks and multimedia. It can even be the electronic reading device such as a Rocket eBook or Pocket PC.
www.planetebook.com/mainpage.asp

A book that is available in electronic format. Usually eBooks are available in Adobe PDF or eBook Reader format, or in Microsoft's LIT format.
www.groundbreaking.com/glossary.htm

an electronically formatted book, designed to be read from a computer, which may be a desktop computer, a handheld computer, or a specialized electronic book reading device.
www.mansfieldct.org/schools/mms/palms/Meet_the_Team/Gloss...

e book - An electronic book, typically delivered over a computer network. E-book capabilities vary by vendor, but most offer hyperlink navigation and keyword searching within the text. Printing and downloading functions are usually available, but are restricted to protect copyrights. ...
library.hsc.unt.edu/researchtools/LibraryInformationTechnology...

e book - a book that has been digitized and is available online. Because e-books usually exist in print and online versions, they are different from websites. The content in e-books has gone through an editorial process that helps to establish and verify its information.
www.socc.edu/library/pgs/databases/glossary-of-research-terms.s...

e book - An electronic version of a traditional print book that can be read by using a personal computer and/or an e-Book reader. McNeese Library offers e-books through its subscription to netLibrary. ...
library.mcneese.edu/tutorial/glossary.htm

What is an ebook?

a) Physical book ⟶ Scanned Page images ⟶ PDF (+OCR)

b) Born digital text   reflowable eBook format
(.doc, .indd, .html, .xml, .txt, .pdf, etc.) ⟶   (EPUB, Mobi/Kindle, HTML, etc.)
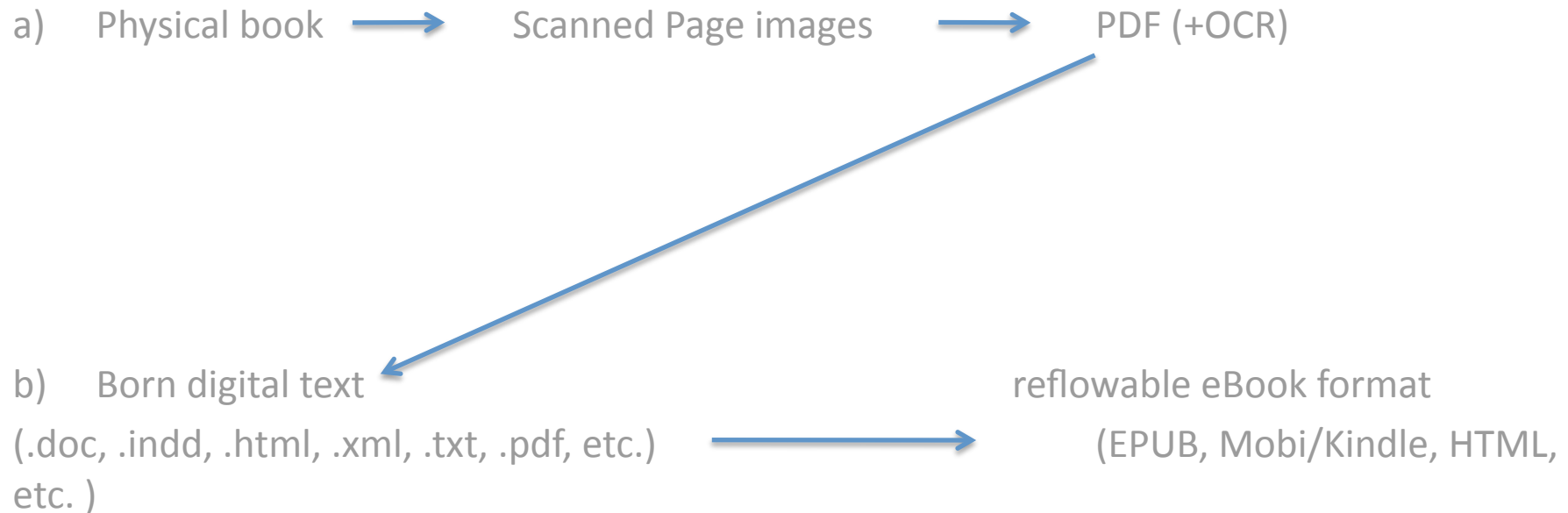
Traditionally, a) might be thought of as the library approach, while b) might be thought of as the publisher approach. But these lines are now crossing, blurring, and disappearing altogether

What formats are out there?

There are more than 30 ebook formats described at
http://en.wikipedia.org/wiki/Comparison_of_e-book_formats

Big players include:

PDF (.pdf)

- Developed by Adobe systems
- Originally intended to reproduce the fixed appearance of a printed page; still typically used for this purpose
- Readable on many devices and platforms

- Kindle (.azw)
  - Proprietary ebook format used by Amazon.
  - Based on the open Mobipocket format.
  - Reflowable text

- EPUB
  - Reflowable text format
  - Open standard used by many devices.
  - Works by combining common formats like .html and .zip in specified ways

EPUB Overview

• Open standard created by the International Digital Publishing Forum (http://idpf.org/)

• Many publishers and vendors offer EPUB for sale or free download (Google Books, Internet Archive, Project Gutenberg, Barnes & Noble...)

• Many devices and apps read it (Nook, Sony Reader, Kobo, iBook, other mobile apps, web browsers with appropriate plug-ins....)

• EPUB 2.0.1 is the latest stable release—from September 2010. EPUB 3.0—with many new features—is under development and expected to be released in 2011

EPUB Overview (continued)

- The EPUB specification relies on three other standards:
    - Open Publication Structure (OPS) : Defines how content is encoded
    http://idpf.org/epub/20/spec/OPS_2.0.1_draft.htm

    - Open Packaging Format (OPF): Defines how all files and metadata are packaged together and relate to one another
    http://idpf.org/epub/20/spec/OPF_2.0.1_draft.htm

    - Open Container Format (OCS): Defines how all files are bundled together
    http://www.idpf.org/doc_library/epub/OCF_2.0.1_draft.doc

The result—EPUB—is a renamed .zip file containing XML and XHTML files (and some other files, like images), that point to one another in standard ways.

Anyone who understands the structure can create or edit EPUB— no proprietary tools or software are required.

The anatomy of EPUB

1. Go to http://www.archive.org/details/luluslibrary00alco and download the EPUB of the book LuLu's Library.

2. EPUB is nothing more than a renamed ZIP archive containing various files. Unzip the EPUB to get at what's inside. If you're having trouble unzipping the file, raise your hand—we'll pass around a set of files already unzipped

3. Once you have a directory named `luluslibrary00alco`, containing directories called `META-INF` and `OEBPS`, we're ready to move on
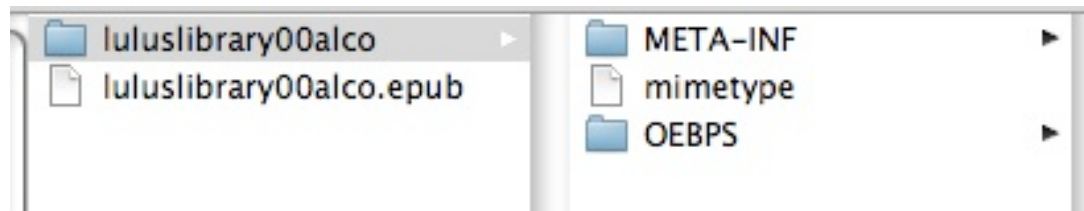
4. If you're waiting for others, you might also go ahead and upload the EPUB (the zipped package) to a reader of your choice

The anatomy of EPUB

An EPUB file contains the following subdirectories and files:

- `mimetype` - tells software that tries to open the file what it is

- `META-INF` directory
    - contains, at minimum, the `container.xml` file, which tells the ereader software where to find the book content.

- `OEBPS` directory - Recommended location for the book's content. It contains:
    - `images` folder - images go here
    - `Content.opf` - XML file that serves as a manifest, or list, of what's here
    - `toc.ncx` – table of contents that allows navigation through the book
    - `.html` files – actual content of your book!

The anatomy of EPUB: `mimetype`

This file is just a plain ASCII text file that contains the line:

`application/EPUB+zip`

The software can look at this file to figure out that this is EPUB.

Try looking at it in a text editor:
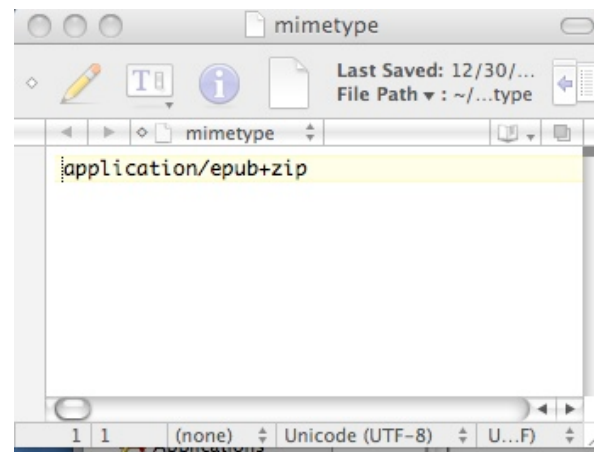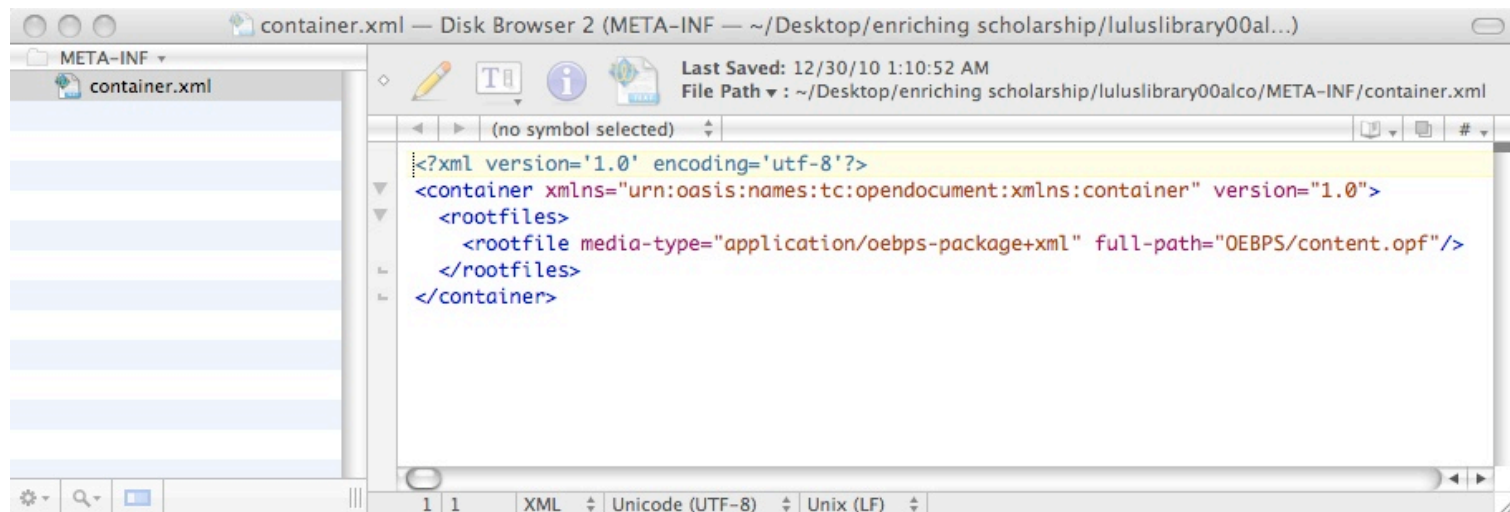
The anatomy of EPUB: `META-INF`

This directory contains the file `container.xml`, which tells the software where to look for a full listing of the content of the book.

M Publishing |University of Michigan Library | Enriching Scholarship | May 4, 2011

The anatomy of EPUB: OEBPS

• IDPF specification recommends (but doesn't require) that your book content goes in OEBPS. Most ereaders expect this. This directory contains the following files and subdirectories:

• `images`	All image files in the book go in this directory
• `Content.opf`	This is an xml file containing Dublin Core metadata about the book and a manifest of all the files contained in the book
• `toc.ncx`	Table of contents for the book
• `xhtml files`	The actual content of the book!
• `css file`	Controls how the book content appears

META-INF ▸	content.opf	cover-image.jpg
mimetype	images ▸	leaf-image0003.jpg
OEBPS ▸	intro.html	leaf-image0005.jpg
	leaf0001.html	leaf-image0006.jpg
	leaf0003.html	leaf-image0011.jpg
	leaf0005.html	leaf-image0294.jpg
	leaf0006.html	picture0.jpg
	leaf0011.html	picture1.jpg
	leaf0294.html	picture2.jpg
	part0000.html	picture3.jpg
	part0001.html	picture4.jpg
	part0002.html	picture5.jpg
	part0003.html	picture6.jpg
	stylesheet.css	picture7.jpg
	toc.ncx	picture8.jpg
		picture9.jpg
		picture10.jpg
		picture11.jpg
		picture12.jpg
		picture13.jpg

Adapted from http://www.jedisaber.com/eBooks/tutorial.asp
Used with the permission of the author

The anatomy of EPUB: `OEBPS/content.opf`

Has three parts:

1. `<metadata>` lists information about the book (author, title, publisher, etc.) using the Dublin Core standard
2. `<manifest>` lists all of the files in book and provides a link to each
3. `<spine>` lists only the text content of the book—the XHTML files—in their proper reading order

```
<package xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns="http://www.idpf.org/2007/opf" version="2.0" unique
  <metadata>
    <meta content="cover-image" name="cover"/>
    <dc:type>Text</dc:type>
    <dc:title>Lulu's library</dc:title>
    <dc:creator>Alcott, Louisa May, 1832-1888</dc:creator>
    <dc:subject>Children's stories, American</dc:subject>
    <dc:description>I. A Christmas dream. The candy country. Naughty Jocko.
    <dc:publisher>Boston, Roberts Brothers</dc:publisher>
    <dc:date>1886-89</dc:date>
    <dc:language>en</dc:language>
    <dc:contributor>New York Public Library</dc:contributor>
    <dc:identifier id="bookid">luluslibrary00alco</dc:identifier>
  </metadata>
```

```
<manifest>
  <item href="stylesheet.css" id="css" media-type="text/css"/>
  <item href="images/cover-image.jpg" id="cover-image" media-type="image/jpeg"/>
  <item href="leaf0001.html" id="leaf0001" media-type="application/xhtml+xml"/>
  <item href="intro.html" id="intro" media-type="application/xhtml+xml"/>
  <item href="images/leaf-image0003.jpg" id="leaf-image0003" media-type="image/jpeg"/>
  <item href="leaf0003.html" id="leaf0003" media-type="application/xhtml+xml"/>
  <item href="images/leaf-image0005.jpg" id="leaf-image0005" media-type="image/jpeg"/>
  <item href="leaf0005.html" id="leaf0005" media-type="application/xhtml+xml"/>
  <item href="images/leaf-image0006.jpg" id="leaf-image0006" media-type="image/jpeg"/>
  <item href="leaf0006.html" id="leaf0006" media-type="application/xhtml+xml"/>
  <item href="part0000.html" id="part0000" media-type="application/xhtml+xml"/>
  <item href="images/leaf-image0011.jpg" id="leaf-image0011" media-type="image/jpeg"/>
  <item href="leaf0011.html" id="leaf0011" media-type="application/xhtml+xml"/>
  <item href="images/picture0.jpg" id="picture0" media-type="image/jpeg"/>
  <item href="images/picture1.jpg" id="picture1" media-type="image/jpeg"/>
  <item href="images/picture2.jpg" id="picture2" media-type="image/jpeg"/>
  <item href="images/picture3.jpg" id="picture3" media-type="image/jpeg"/>
  <item href="images/picture4.jpg" id="picture4" media-type="image/jpeg"/>
  <item href="images/picture5.jpg" id="picture5" media-type="image/jpeg"/>
  <item href="images/picture6.jpg" id="picture6" media-type="image/jpeg"/>
  <item href="images/picture7.jpg" id="picture7" media-type="image/jpeg"/>
  <item href="images/picture8.jpg" id="picture8" media-type="image/jpeg"/>
  <item href="images/picture9.jpg" id="picture9" media-type="image/jpeg"/>
```

```
<spine toc="ncx">
  <itemref idref="leaf0001" linear="no"/>
  <itemref idref="intro" linear="no"/>
  <itemref idref="leaf0003" linear="no"/>
  <itemref idref="leaf0005" linear="no"/>
  <itemref idref="leaf0006" linear="no"/>
  <itemref idref="part0000"/>
  <itemref idref="leaf0011" linear="no"/>
  <itemref idref="part0001"/>
  <itemref idref="part0002"/>
  <itemref idref="part0003"/>
  <itemref idref="leaf0294" linear="no"/>
</spine>
```

Adapted from http://www.jedisaber.com/eBooks/tutorial.asp
Used with the permission of the author

The anatomy of EPUB: `toc.ncx`

NCX= Navigation Control for XML applications

Controls how your EPUB reader will render the book's table of contents by documenting for each section of the book its order, unique ID, level in the hierarchy of the book, label, and where the file containing the content is

```
<navPoint playOrder="1" id="navpointnavpoint000001" class="navpoint-level1">
  <navLabel>
    <text>Front Cover</text>
  </navLabel>
  <content src="leaf0001.html"/>
</navPoint>
<navPoint playOrder="2" id="navpointnavpoint000002" class="navpoint-level1">
  <navLabel>
    <text>Title Page</text>
  </navLabel>
  <content src="leaf0005.html"/>
</navPoint>
```

Lulu's library

Alcott, Louisa May,
1832–1888

Front Cover

Title Page

Copyright

Pages

Table of Contents

Back Cover

Download this book

Book information

LULU'S LIBI

The anatomy of EPUB: Content!

```html
<head>
  <title>part0001</title>
  <meta content="abbyy to epub tool, v0.2" name="generator"/>
  <link href="stylesheet.css" type="text/css" rel="stylesheet"/>
  <meta content="application/xhtml+xml; charset=utf-8"
  http-equiv="Content-Type"/>
</head>
<body>
  <div class="body">
    <p class="illus">
      <img src="images/picture0.jpg" alt="picture0"/>
    </p>
    <p> Instead of dying in her cell, the fairy had made it beautiful. -  PAGE
    13.</p>
    <p> I.</p>
    <p> THE FROST KING AND HOW THE FAIRIES CONQUERED  HIM.</p>
    <p> r   I ^HE Queen sat upon her throne, and all the fairies from the four
    kingdoms were gathered for a grand council. A very important question was
    to be decided, and the bravest, wisest elves were met to see what could be
    done.</p>
    <div class="newpage" id="page-8">
    <p> The Frost King made war upon the flowers; and it was a great grief to
    Queen Blossom and her subjects to see their darlings die year after year,
    instead of enjoying one long summer, as they might have done but for him.
    She had sent messengers with splendid gifts, and had begged him to stop
    this dreadful war, which made autumn so sad and left the fields strewn
    with dead flowers. But he sent back the gifts, sternly refused her
    prayers, and went on with his cruel work; because he was a tyrant, and
    loved to destroy innocent things.</p>
    <p> " My subjects, we will try once more," said the Queen, " if any one
    can propose a plan that will touch his hard heart and make him kind to the
    dear flowers."</p>
```

LULU'S LIBRARY

Instead of dying in her cell, the fairy had made it beautiful. — PAGE 13.

I.

THE FROST KING AND HOW THE FAIRIES CONQUERED HIM.

r I ^HE Queen sat upon her throne, and all the fairies from the four kingdoms were gathered for a grand council. A very important question was to be decided, and the bravest, wisest elves were met to see what could be done.

The Frost King made war upon the flowers; and it was a great grief to Queen Blossom and her subjects to see their darlings die year after year, instead of enjoying one long summer, as they might have done but for him. She had sent messengers with splendid gifts, and had begged him to stop this dreadful war, which made autumn so sad and left the fields strewn with dead flowers. But he sent back the gifts, sternly refused her prayers, and went on with his cruel work; because he was a tyrant, and loved to destroy innocent things.

" My subjects, we will try once more," said the Queen, " if any one can propose a plan that will touch his hard heart and make him kind to the dear flowers."

The Anatomy of EPUB: Putting it All Together

`Mimetype`

hey, this is epub!

`META-INF/content.xml`

Aha! I need to go to
`OEBPS/content.opf`

`OEBPS/content.opf`

Everything here? Good! Let's read!

Ereader
software

`toc.ncx` → `*.html`

`images/*.jpg`
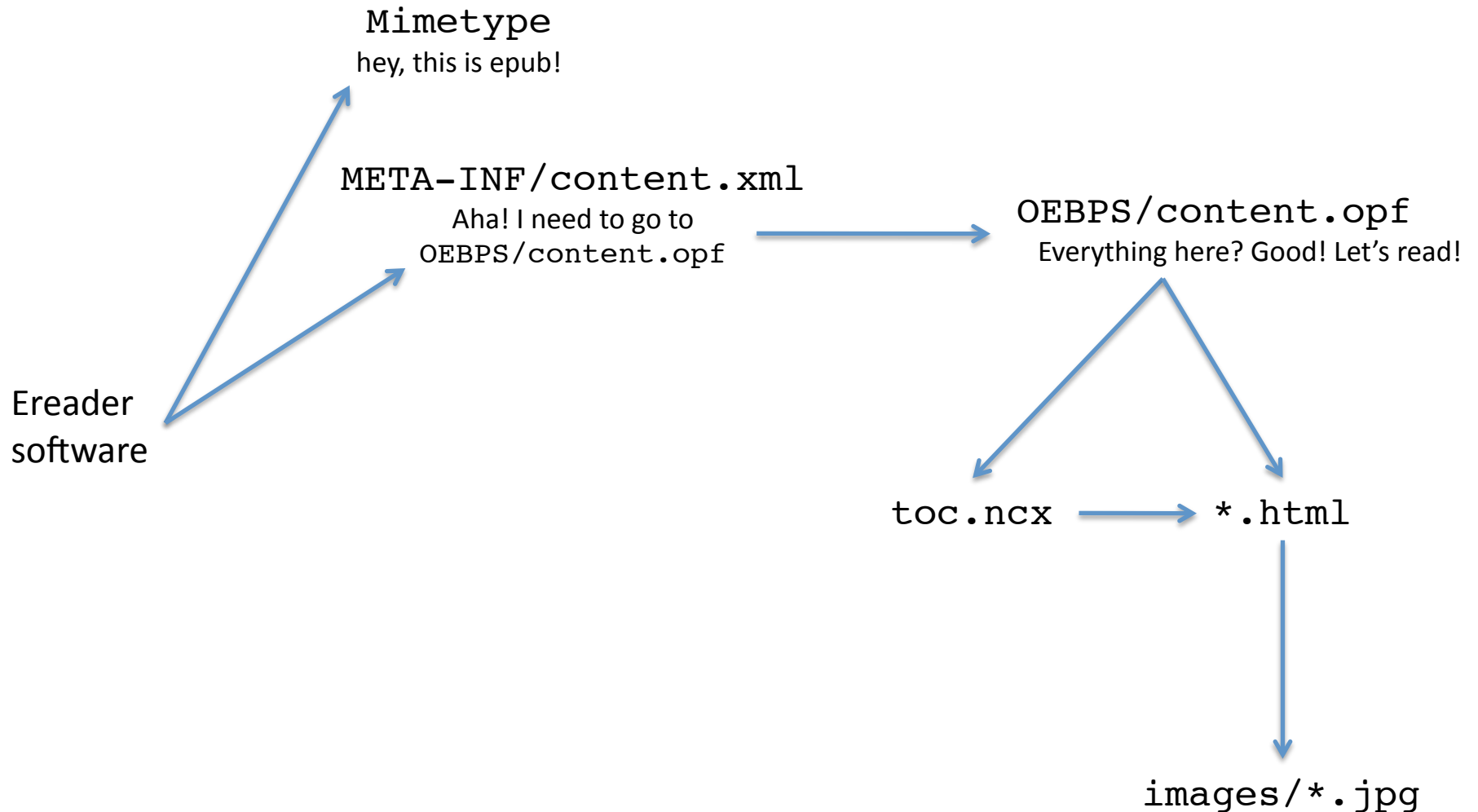
But how do I *make* EPUB from my own work?

On making EPUB

You don't have to do it alone!

← ───────────────────────────── →

Create ebooks by hand

• Free
• No special software required
• Online tutorials can help
• If your book isn't already encoded in decent html, this is going to be a lot of work.

Use a tool to convert Edit/tweak by hand

• Lightens workload
• Possible to convert from many formats
• Full control over file
• Still need to get book into distribution channels, if desired

Use a service to make and distribute your ebooks

• Cost may vary
• Control may vary
• Less work
• Gets book into various sales/ distribution channels

Full service ebook conversion, distribution

- Services such as LuLu [http://www.lulu.com/] (which feeds into the Apple iBookstore) and PubIt! [http://pubit.barnesandnoble.com/] (Barnes & Noble sales channels) can provide help with converting your work to EPUB and make it available for sale

- Amazon's Kindle Direct Self Publishing [http://kdp.amazon.com/] does the same for the Kindle format and the Amazon sales stream

Tools to convert

Many free tools automatically create EPUB from other file formats

OxGarage [http://www.tei-c.org/ege-webclient/]

2EPUB [http://www.2epub.com/]

Calibre [http://calibre-ebook.com/]

…and many more!

But automated conversion will only get you so far…

Try it! (1)

1. Download our guinea pig files: my undergraduate senior thesis (as .doc, .pdf, or both) [http://www.lib.umich.edu/digital-publishing-production/care-and-keeping-of-ebooks-may-4-2011]

2. Try converting to EPUB using one or more of these tools. Do at least two different conversions

3. Upload the resulting EPUBs to a reader. Differences between the two are to be expected—each tool will map the source file to EPUB a bit differently, and the PDF and Word doc will each produce different results.

4. What changes would you like to see for this ebook to correctly represent the original paper? What parts of the EPUB package need to be edited?

5. Challenge: Unzip the EPUB and use your text editor to look at the files. Try to pinpoint places where changes need to be made, such as to the `toc.ncx`, the metadata, or to formatting in the .html files. (You don't need to know how to make the changes!)

Try it! (2)

Two ways to change the EPUB:

- Unzip the EPUB and edit it directly. This way, you'll know exactly what you've changed and have more control over the output

- Experiment with formatting in the Word document, run it through the conversion again, and look at the EPUB output in a reader. Does the output change if you apply styles to section headings? If you insert a dynamic Table of Contents in the Word document?

- Try it with your own documents; try converting from other formats like .html

Zipping up your EPUB

If you edit the EPUB directly, you'll need to re-zip the files before you upload them to a reader. There's a specific way this needs to be done so that the mimetype file doesn't get compressed and appears in the right place. Tips below come from Aaron DeMott:

Windows:
- Create an empty .zip file with whatever name you like (See notes below for creating the container in OS X)
- Copy the mimetype file into the zip file (don't use compression on this file)
- Copy the rest of the files and folders into the zip file
- Re-name the .zip extension to .epub
- NB: the EPUB specification requires the mimetype file to be the first file in the container. If you are using a GUI zip program, make a blank ZIP file, then copy just the mimetype file into the zip file, followed by the rest of the files. If you are using a command line zip utility, then either add the mimetype file first, or if your zip program of choice supports adding multiple files/folders at a time, make sure the mimetype file is the first one in the list.

Mac OS X
- Prepare your content like normal
- Right-click the folder the files are in (ensure they are in their own file first) and click "Compress Folder"
- Change the "zip" extension to "epub"
- The problem with this method is that the mimetype file isn't necessarily added to the zip folder first. A user at the mobileread forums has created an Applescript that will do the zipping for you, in the right order. You can get it here: http://www.mobileread.com/forums/showthread.php?t=55681

Adapted from http://www.jedisaber.com/eBooks/tutorial.asp
Used with the permission of the author

Take Aways?

- EPUB is not a "black box." It's a structured set of files that you can control (barring DRM)

- You can build EPUB yourself from the ground up, get a service to do it for you, or something in between

- Many tools are available to package your text into EPUB—but some manual editing, requiring familiarity with EPUB and HTML, will be needed for the book to look right.

- Formatting makes a difference! Submitting highly consistent, structured electronic text to the conversion tool will help your EPUB output from the get-go

Questions? Ideas? Discussion?

Please fill out your evaluation!

http://tinyurl.com/myES2011