

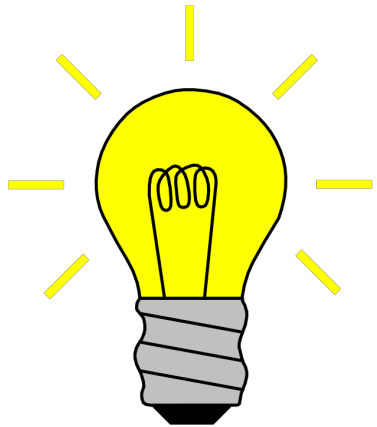
# **The Case for Authoring and Producing Books in (X)HTML5**

**Balisage 2013**  
**August 7, 2013**

Sanders Kleinfeld  
*O'Reilly Media, Inc.*

**The Goal of  
Publishing:**

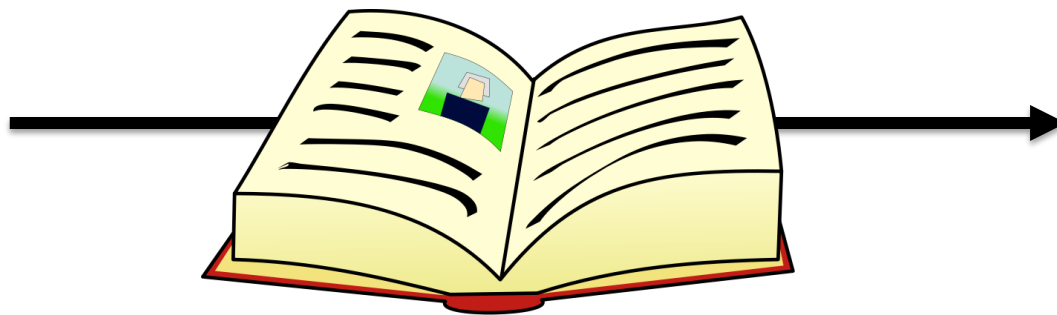
**Packaging and  
Distribution of Ideas**



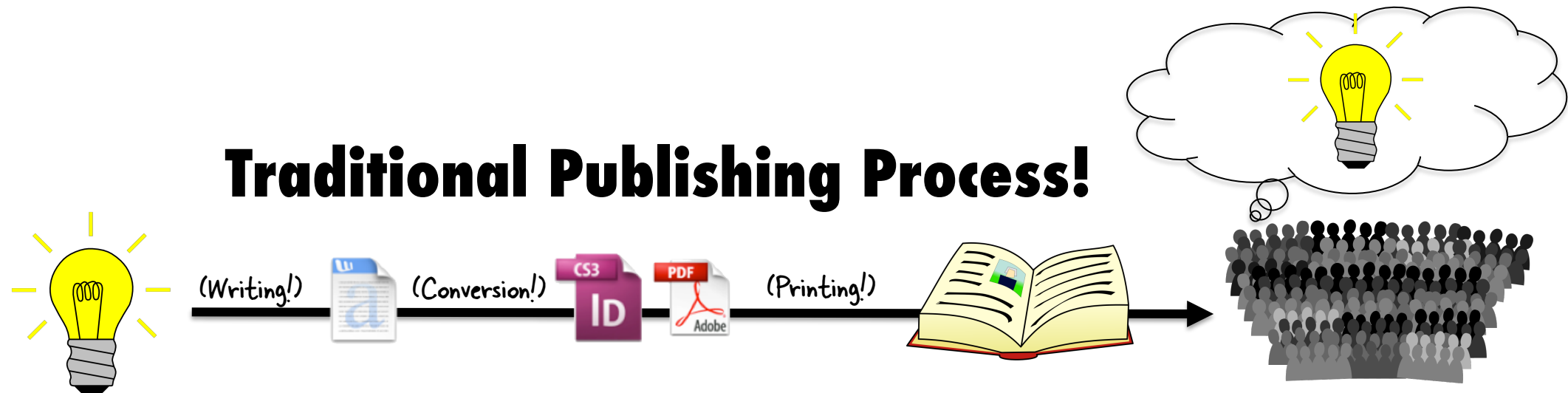
**Publishing!**



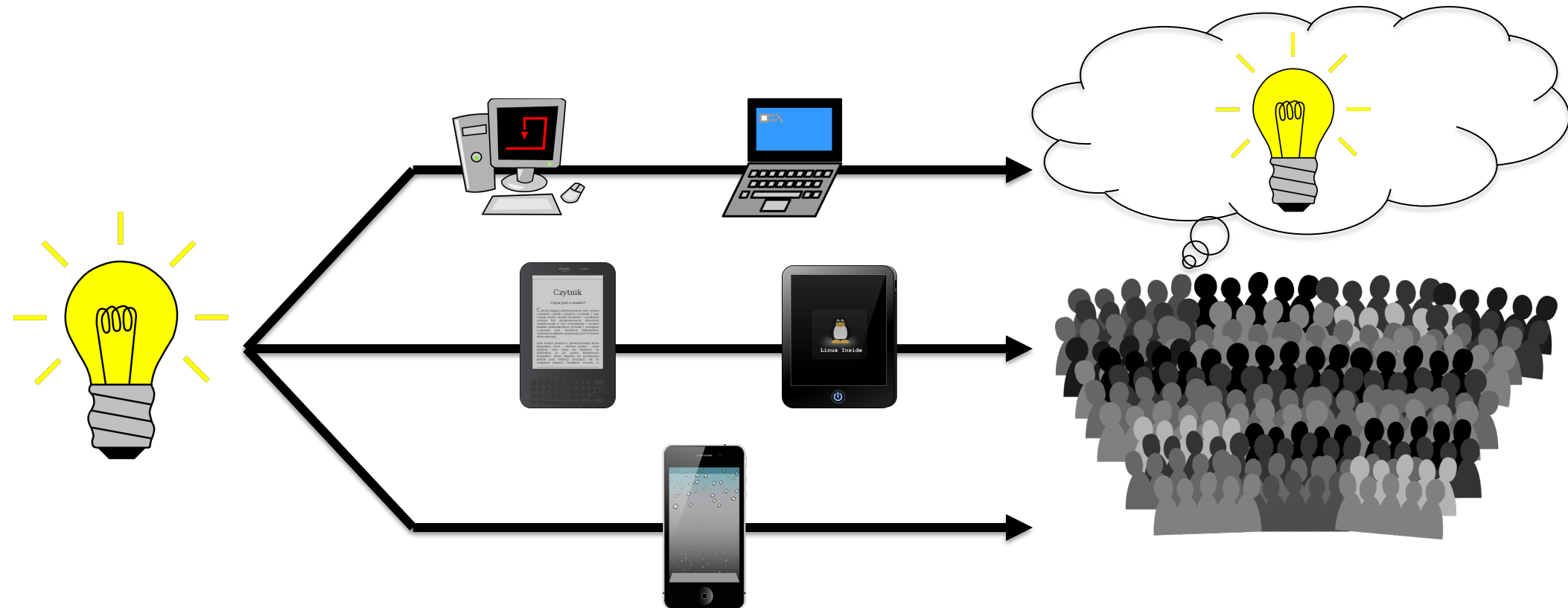
# **Traditional Publishing!**



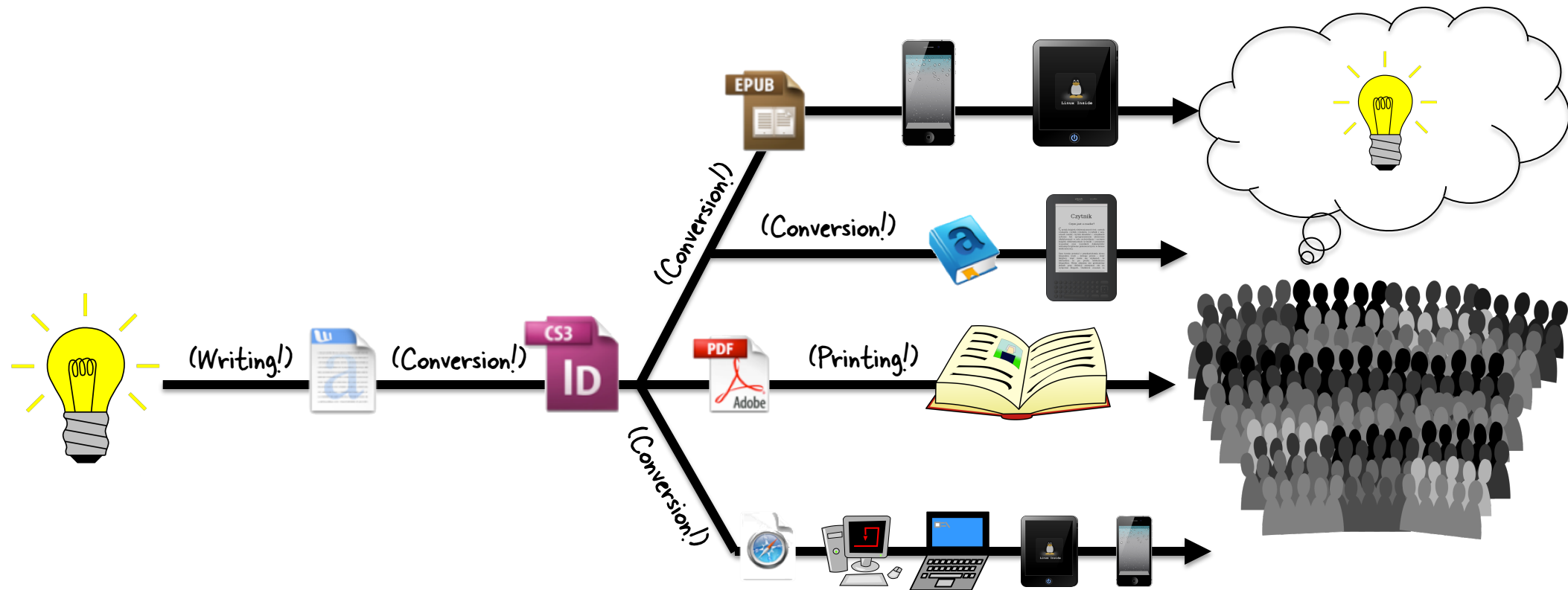
# Traditional Publishing Process!



# Digital Publishing!

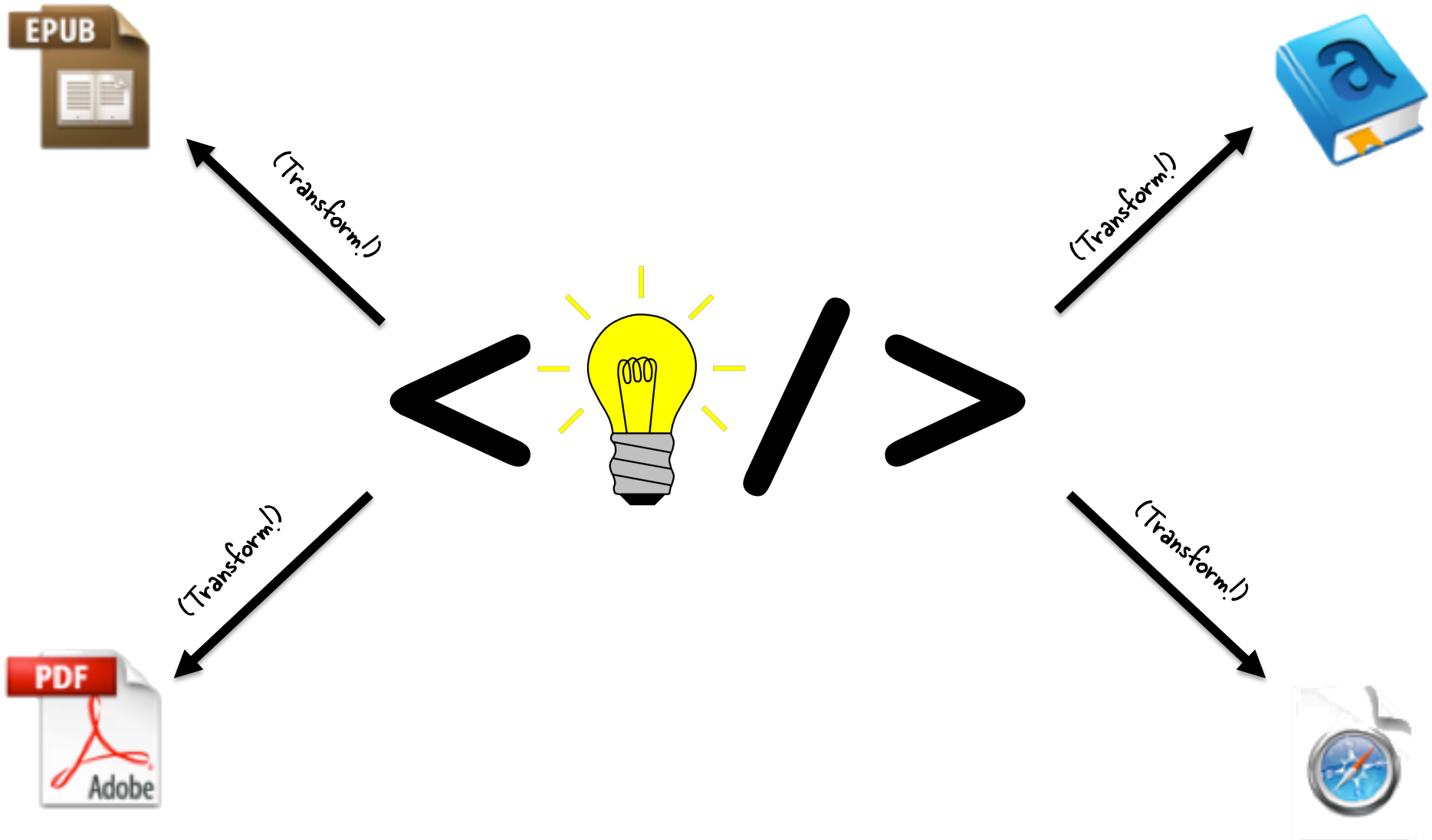


# (Post)Modern Publishing Process! (Both Print and Digital)



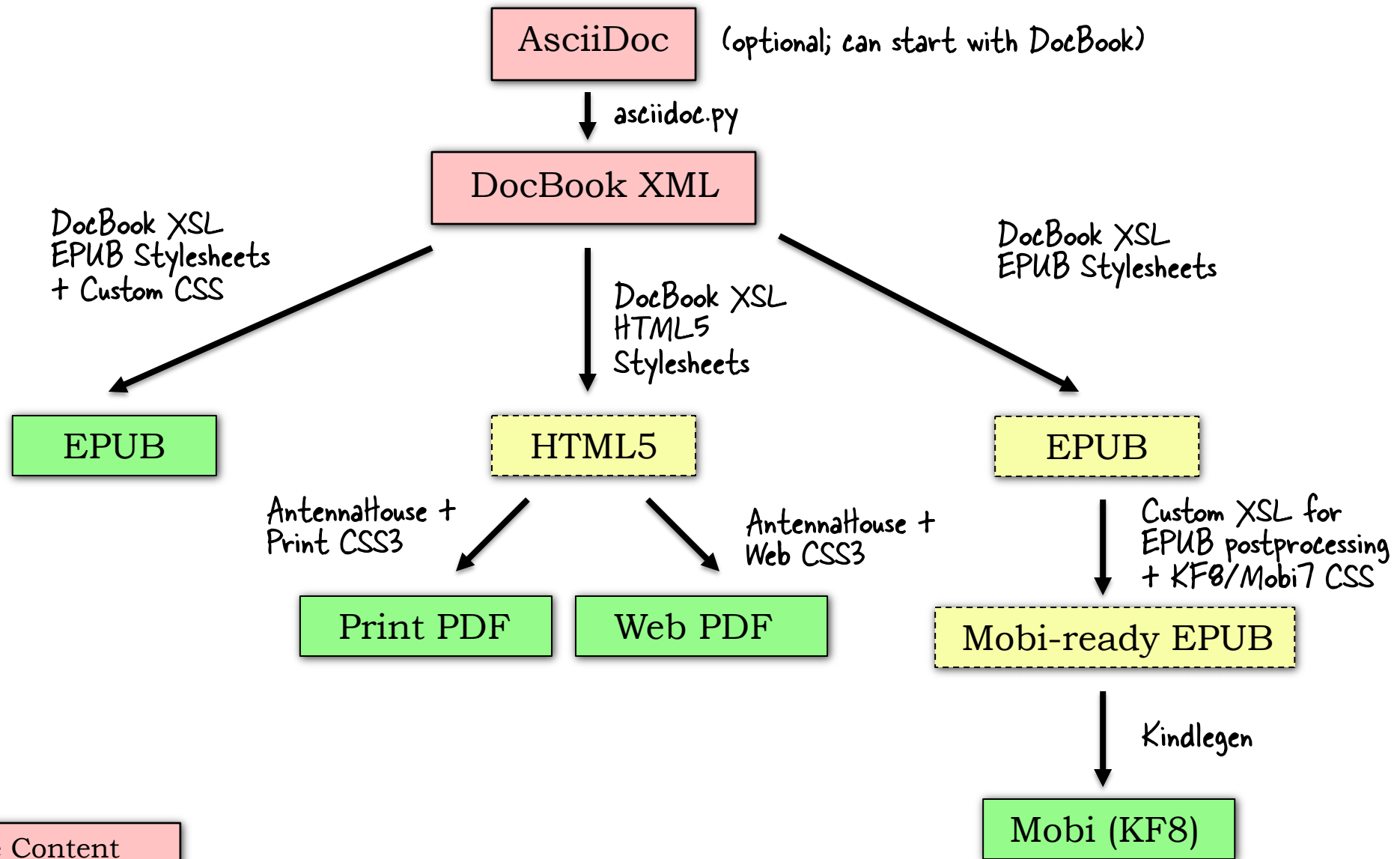
# The Single-Source Solution:

**Replace conversions with semantic markup and automated transforms**





# ASC/DB Single-Source Workflow:



Source Content

Intermediate Output

Final Output For Sale

# **Three Realizations About a DocBook-Source Workflow**

**Realization #1:  
DocBook XSL  
toolchain is rather  
heavyweight,  
complex**

# DocBook XSL Toolchain Stats\*

**The XHTML5 stylesheets contain:**

- **33,707 lines of HTML-generation code...**
- **...which rely on 8,346 lines of common dependencies**

**Or, in terms of functions, they contain:**

- **1,857 <xsl:template>s...**
- **...which rely on 272 common dependency <xsl:template>s**

# Customization Use Case:

Find the code that generates XHTML title page content for chapters, and change it to use `<h2>` instead of `<h1>` for title heading.

# Customization Use Case: Step by Step

1. Start by finding `<xsl:template match="chapter">` in *component.xsl*, which calls...
2. `<xsl:template name="chapter.titlepage">` in *titlepage.templates.xsl*, which in turn calls...
3. `<xsl:template name="chapter.titlepage.before.recto">`,  
`<xsl:template name="chapter.titlepage.recto">`,  
`<xsl:template name="chapter.titlepage.before.verso">`, and  
`<xsl:template name="chapter.titlepage.verso">`, which in turn call...\*

\* chapter.titlepage.before.recto, chapter.titlepage.before.verso, and chapter.titlepage.verso are empty in docbook-epub3-addon-b3 distro

# Customization Use Case: Step by Step (*continued*)

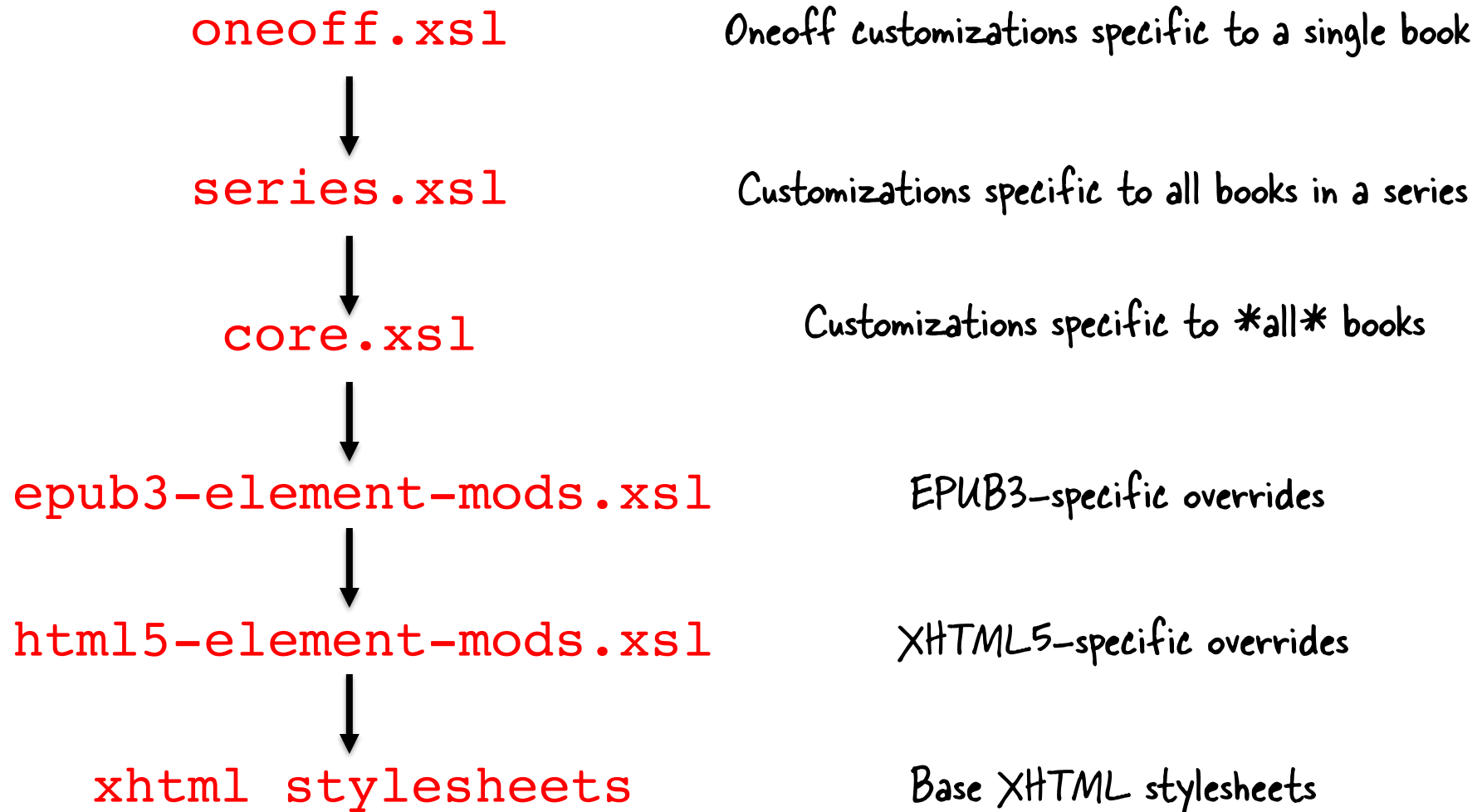
4. `<xsl:template>` matches on the elements `chapterinfo/title`, `docinfo/title`, `info/title`, `title`, `chapterinfo/subtitle`, `docinfo/subtitle`, `info/subtitle`, and `subtitle` in **`mode="chapter.titlepage.recto.auto.mode"`**, which in turn call...
5. `<xsl:template>` matches on the elements `chapterinfo/title`, `docinfo/title`, `info/title`, `title`, `chapterinfo/subtitle`, `docinfo/subtitle`, `info/subtitle`, and `subtitle` in **`mode="chapter.titlepage.recto.mode"`**, which in turn call...

# Customization Use Case: Step by Step (*continued*)

6. `<xsl:template>` matches on the elements `chapterinfo/title`, `docinfo/title`, `info/title`, `title`, `chapterinfo/subtitle`, `docinfo/subtitle`, `info/subtitle`, and `subtitle` in **mode="titlepage.mode"** in *titlepage.xsl* which in turn contain...
7. The logic we're looking for that outputs an `<h1>` element for the chapter title. Hooray! We need to override `<xsl:template match="title" mode="titlepage.mode">`



# Customizations in docbook-xsl: Five Levels of EPUB Overrides



# Hierarchy of overrides =



Photographer: Derek Mawhinney on Oct. 23, 2005 Subject: A game of [Jenga](r) in progress  
Cafe Nervosa 21:19, 31 October 2005 (UTC)-- <http://en.wikipedia.org/wiki/File:Jenga.JPG>

**This complexity is  
a necessary evil,  
emphasis on *evil***

# **Realization #2: DocBook Not Ideal for Digital-First Content**

# Digital-First Content Development

When doing digital-first (ebook/web) content development, these are the key output formats:

- **EPUB (2.0 and 3.0)**



- **Amazon Kindle Mobi (Mobi7/KF8)**



- **PDF**



- **HTML**



# The Common Thread: HTML + CSS



= HTML + CSS + open source packaging



= HTML + CSS + proprietary packaging



= HTML + CSS + PDF processor\*



= HTML + CSS (duh!)

\* e.g., AntennaHouse Formatter or Prince

# Interactivity/Multimedia Is Ultimately All About HTML5

`<canvas>`

`<audio>`

`<video>`

`<math>`

`<svg>`

**Realization #3:  
Not All Authors  
Love Writing in  
DocBook ☹️**



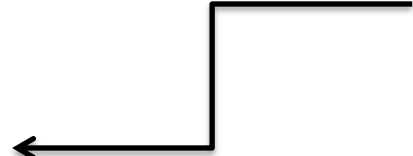
**Authors prefer  
WYSIWYG authoring  
platforms because...**

**“Nobody’s going to  
learn your markup  
language”**

Books in Browsers 2012: Liza Daly & Keith Fahlgren,  
“The self-publishing book”  
<http://www.youtube.com/watch?v=UWftLHopWQ0#t=5m25s>

# Non-Technical Authors Don't Like This...

DocBook



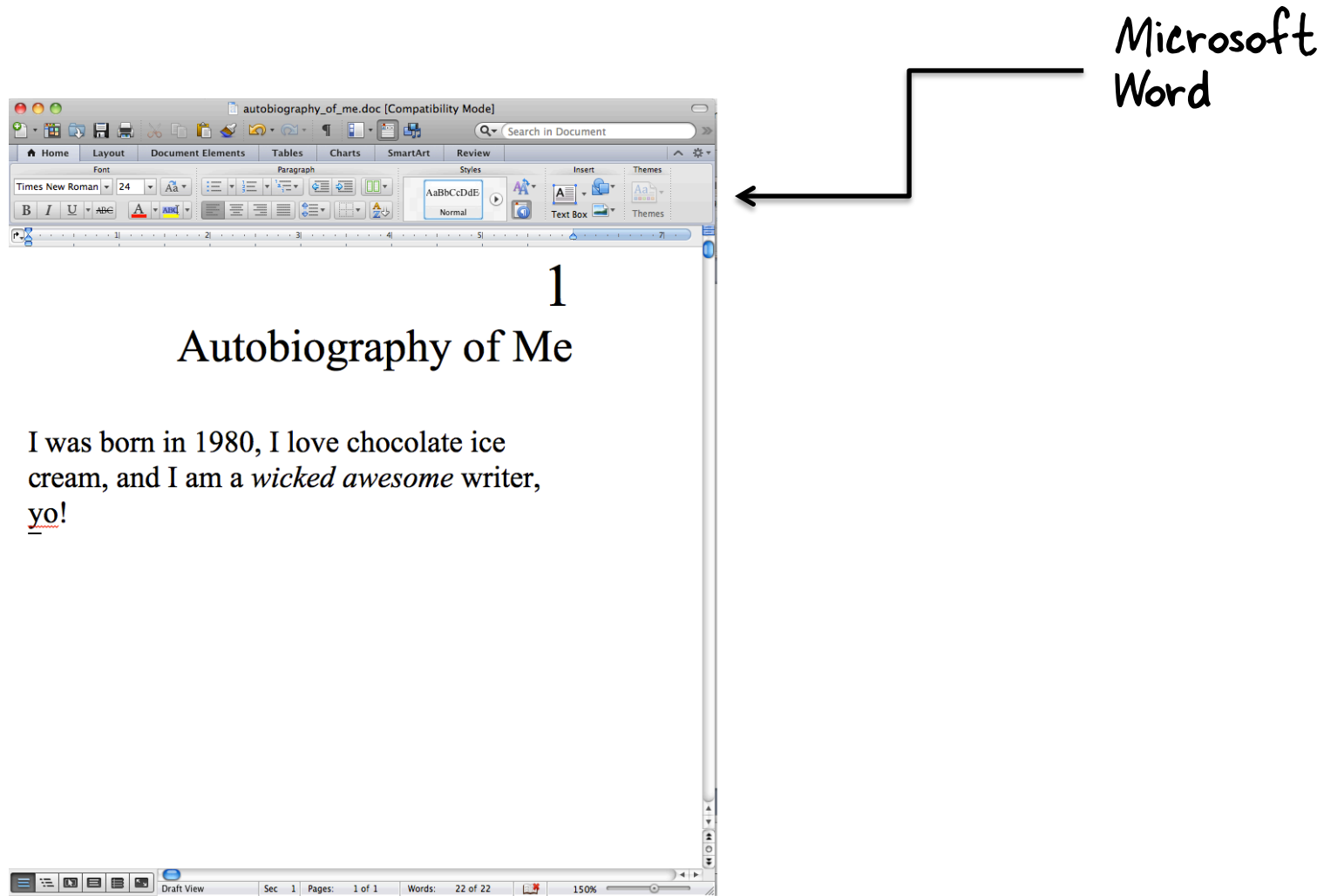
```
<?xml version="1.0" encoding="utf-8"?>  
<!DOCTYPE chapter PUBLIC "-//OASIS//DTD DocBook XML V4.5//EN"  
"http://www.oasis-open.org/docbook/xml/4.5/docbookx.dtd">  
<chapter>  
  <title>Autobiography of Me</title>  
  <para>I was born in 1980, I love chocolate ice cream, and I  
am a <emphasis>wicked awesome</emphasis> writer, yo!</para>  
</chapter>
```

# Non-Technical Authors Will Sometimes Tolerate This...

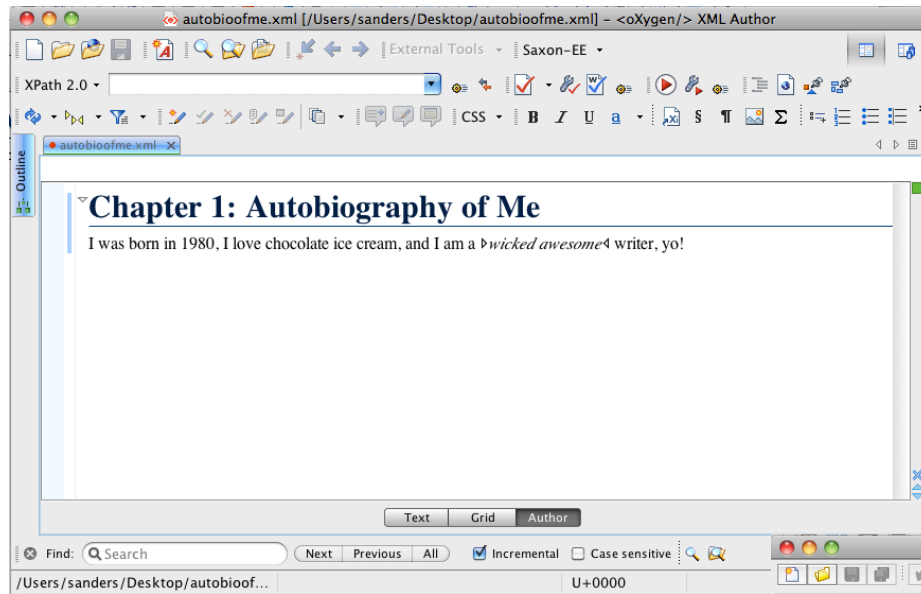
`== Autobiography of Me`  *AsciiDoc*

I was born in 1980, I love chocolate ice cream,  
and I am a wicked awesome writer, yo!

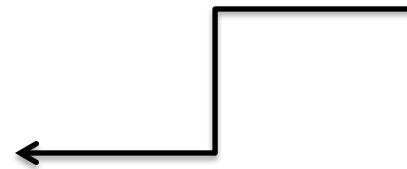
# But Non-Technical Authors Really Want This...



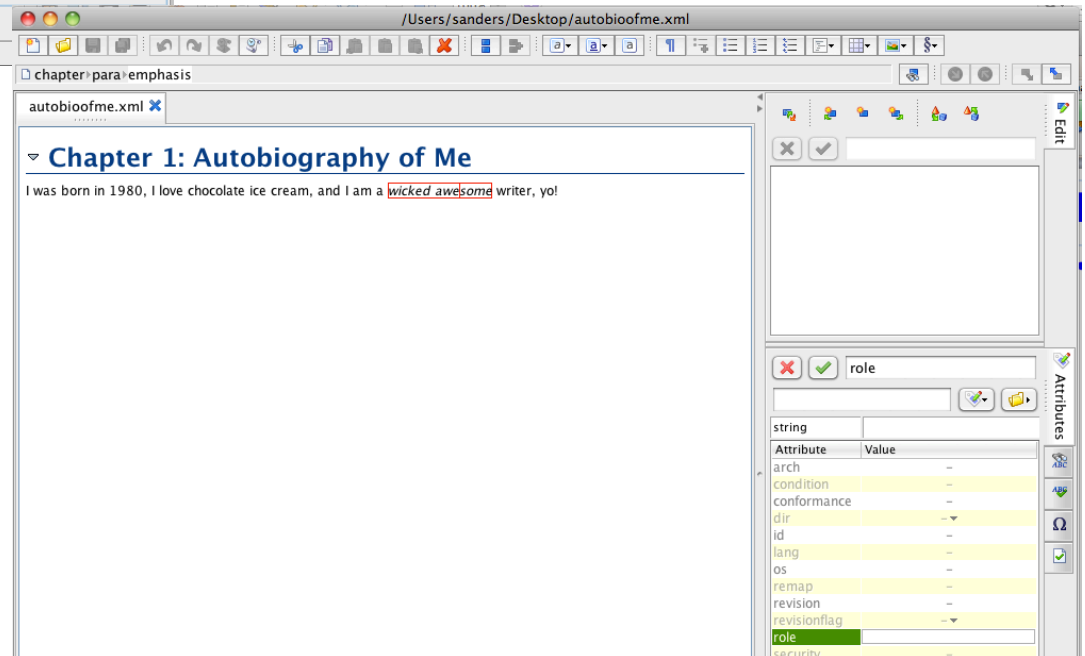
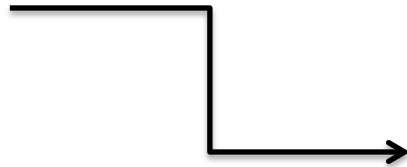
# oXygen and XMLMind Do Make DocBook Authorship More Accessible, but...



oXygen

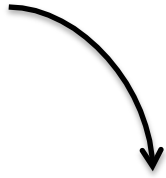


XMLMind



**This is the Future of  
Digital Content  
Creation:**

Medium



(Short-Form Web Publishing)

The screenshot shows a web browser window with the title 'Editing Autobiography of Me'. The address bar shows 'https://medium.com/...' and the search bar contains 'Google'. The browser's bookmark bar includes links to Apple, Yahoo!, Google Maps, YouTube, Wikipedia, News (529), and Popular. The Medium interface features a black 'M' logo and a 'SAVED' status. On the right, there are three buttons: 'Delete', 'Share Draft', and a green 'Publish' button. The main content area has a placeholder image icon, followed by the title 'Autobiography of Me' in a large, bold font. Below the title is a subtitle prompt 'Type your subtitle (optional)'. The body text reads: 'I was born in 1980, I love chocolate ice cream, and I am a wicked awesome writer, yo!'.

Editing Autobiography of Me


https://medium.com/...

Google

Apple Yahoo! Google Maps YouTube Wikipedia News (529) Popular

M SAVED

Delete Share Draft Publish



# Autobiography of Me

Type your subtitle (optional)

I was born in 1980, I love chocolate ice cream, and I am a wicked awesome writer, yo!



O'Reilly  
Atlas

(Short and Long-Form Print,  
Digital, and Web Publishing)

The screenshot shows a web browser window titled "O'Reilly Atlas" with the URL <http://orm-atlas-staging.herokuapp.com/sandersk/Autobioofme/editor/master/autobio.html#visual>. The browser's address bar and search bar are visible. Below the browser window is the O'Reilly Atlas web editor interface. On the left is a sidebar with a red header "Back to Project" and a list of editors: "Choose Editor" (with an eye icon), "Code Editor", "Visual Editor" (selected with a blue dot), and "Files" (with a book icon). Below the list are buttons for "Upload", "Create File", "Create Folder", and "Delete Files". At the bottom of the sidebar is a file list showing "autobio.html" with a red dot. The main editing area has a top toolbar with buttons for bold (B), italic (I), underline (U), subscript (x<sub>2</sub>), superscript (x<sup>2</sup>), strikethrough (I<sub>x</sub>), bulleted list, numbered list, link, unlink, quote, and YouTube. Below the toolbar are two dropdown menus: "Chapter" and "Normal". The main content area is titled "Chapter" and contains the text "Autobiography of me" in a large, bold font. Below the title is a paragraph of text: "I was born in 1980, I love chocolate ice cream, and I am a wicked awesome writer, yo!". The word "awesome" is highlighted in green. On the right side of the interface is a sidebar with a red header and buttons for "Save", "Build" (with a book icon), "PDF", "EPUB", "MOBI", "HTML", and a large blue "Build" button at the bottom.

# **Next-Generation Content Authoring =**

- **WYSIWYG**
- **Web-Based (Responsive Design)**
- **Version-Controlled**
- **Seamless**

**All Roads Lead to  
(X)HTML5...**

**...But Is It Semantic  
Enough for Book  
Publishing?**

oreillymedia/HTMLBook

GitHub, Inc. Reader

Google

Apple Yahoo! Google Maps YouTube Wikipedia News (530) Popular

This repository Search or type a command Explore Gist Blog Help

sandersk

oreillymedia / HTMLBook

Unwatch 15 Star 9 Fork 5

Let's write books in HTML! — Edit

429 commits 12 branches 0 releases 4 contributors

branch: master HTMLBook

Merge pull request #60 from oreillymedia:pull/60/head

sandersk authored 5 hours ago

- htmlbook-xsl Actually generation order does not matter here, as logic is now written. 5 hours ago
- samples Updated files for HTML5. 23 days ago
- schema Tweaked schema input for XML. It was a bit hard to make only happen. 3 months ago
- stylesheets a base stylesheet 3 months ago
- LICENSE adding MIT license 4 months ago
- README.asciidoc Update README.asciidoc 3 months ago
- poster.png Added poster.png 23 days ago
- specification.asciidoc Updated specification to reflect the use of @data-type for formal sem... a month ago

Issues 3

Pull Requests 0

Wiki

Pulse

Graphs

Network

SSH clone URL

git@github.com:or

You can clone with [HTTPS](#), [SSH](#), [Subversion](#), and other methods.

Clone in Desktop

Download ZIP

README.asciidoc

## HTMLBook

Let's write books in HTML! HTMLBook is an open, HTML5-based standard for the authoring and production of both print and digital books. HTMLBook is built on the following premises:

# **HTMLBook =**

- **Open Spec for Book Authoring**
- **Subsets XHTML5 Vocabulary and Content Model**
- **Adds Structural Semantics à la DocBook**
- **Open Source Tooling for Producing Ebook Outputs**

# HTMLBook Schema

```
<!-- Chapter/Appendix/Preface/etc. -->
<xs:complexType name="bookmaindiv">
  <xs:sequence>
    <xs:element ref="h1"/>
    <xs:element name="h2" minOccurs="0"
maxOccurs="unbounded" type="subheading"/>
    <xs:group ref="blockelements" minOccurs="0"
maxOccurs="unbounded"/>
    <xs:element name="section" minOccurs="0"
maxOccurs="unbounded" type="sect1"/>
  </xs:sequence>
  <!-- Note: special data-type attr restrictions here -->
  <xs:attribute name="data-type" use="required"
type="bookmaindivtype"/>
  <xs:attributeGroup ref="globals"/>
</xs:complexType>
```

**([github.com/oreillymedia/HTMLBook/blob/master/schema/htmlbook.xsd](https://github.com/oreillymedia/HTMLBook/blob/master/schema/htmlbook.xsd))**

# HTMLBook Sample

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>HTMLBook Sample</title>
  </head>
  <body data-type="book" id="htmlbook">
    <section data-type="chapter" id="chapter01">
      <h1>Chapter 1. HTMLBook Markup</h1>
      <p>This chapter describes and demonstrates
the types of markup<a data-type="indexterm" data-
primary="markup" data-secondary="types of"></a> that
might appear in a chapter. See <em>mappings.asciidoc</
em> for more information. HTMLBook borrows much of its
semantics from the EPUB 3 specification, as applied
via the <a href="http://idpf.org/accessibility/
guidelines/content/semantics/epub-
type.php"><code>epub:type</code></a> attribute.</p>
    </section>
  </body>
</html>
```

**(github.com/oreillymedia/HTMLBook/blob/master/samples/  
htmlbook.html**



# HTMLBook XSL

**epub.xsl**

**ncx.xsl**

**opf.xsl**

**chunk.xsl**

**htmlbook.xsl**

**common.xsl**

**elements.xsl**

**indexgen.xsl**

**param.xsl**

**tocgen.xsl**

**xrefgen.xsl**

**([github.com/oreillymedia/HTMLBook/tree/master/htmlbook-xsl](https://github.com/oreillymedia/HTMLBook/tree/master/htmlbook-xsl))**

# HTMLBook XSL Toolchain Stats

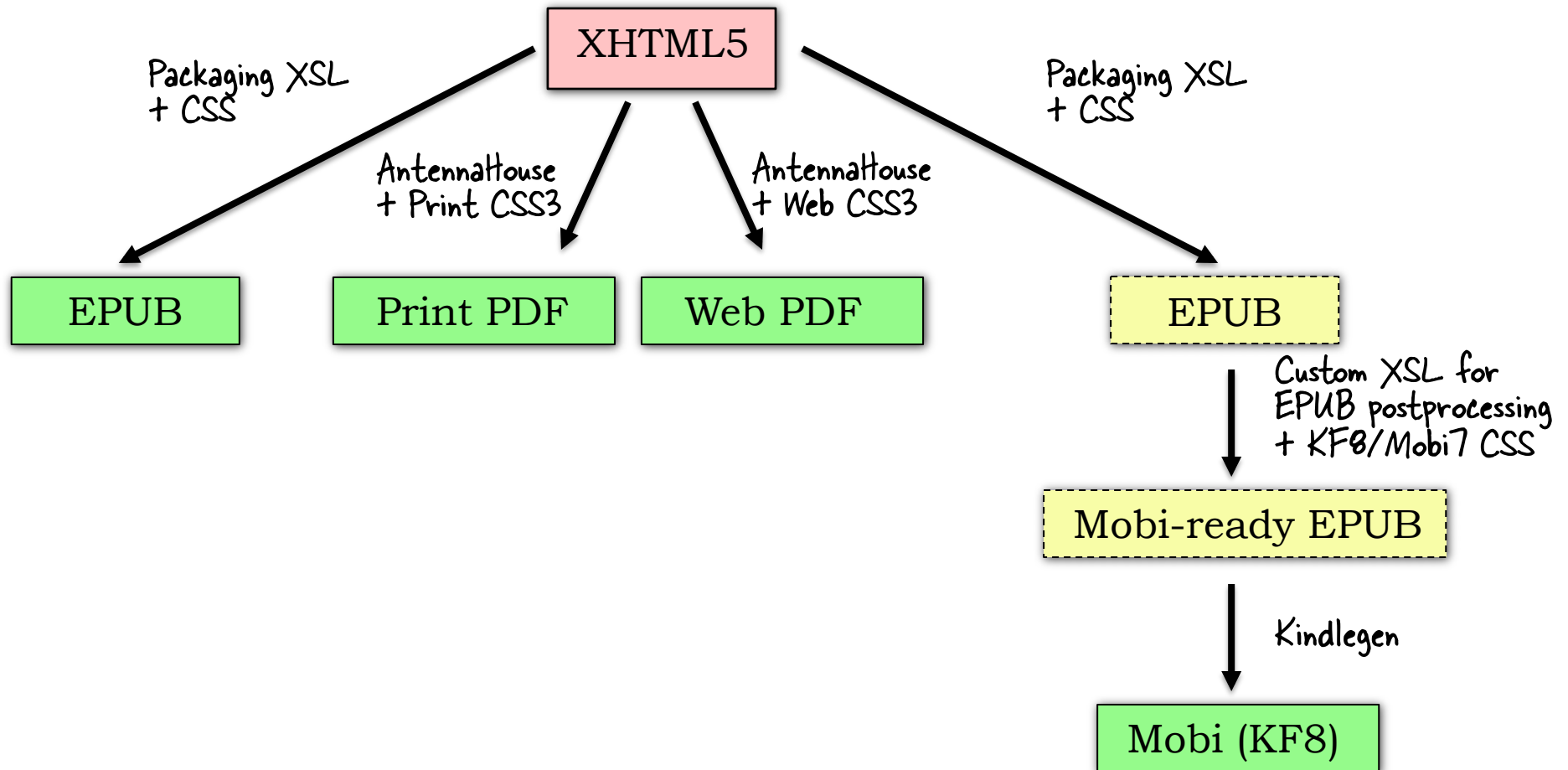
**The HTMLBook stylesheets contain:**

- **2,982 lines of HTML-processing code...**  
**(91% reduction from docbook-xsl)**
- **...which rely on 406 lines of common dependencies**

**Or, in terms of functions, they contain:**

- **73 <xsl:template>s...**  
**(96% reduction from docbook-xsl)**
- **...which rely on 16 common dependency <xsl:template>s**

# XHTML5 Single-Source Workflow:



Source Content

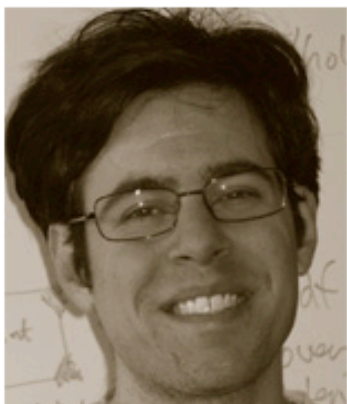
Intermediate Output

Final Output For Sale

# Contact Me!

**Email:** [sanders@oreilly.com](mailto:sanders@oreilly.com)

**Twitter:** [@sandersk](https://twitter.com/sandersk)



## Sanders Kleinfeld

[sanders@oreilly.com](mailto:sanders@oreilly.com)

<http://twitter.com/sandersk>

Cambridge, Massachusetts

*Ebook Alchemist*

### Areas of Expertise:

- DocBook XML
- XSLT
- XPath
- XQuery
- RDF
- SPARQL
- EPUB
- Mobi
- HTML5
- speaking
- writing

#### Biography

#### Books

#### Blog

#### Multimedia

Sanders Kleinfeld has been employed at O'Reilly Media since 2004 and has held a variety of positions, including roles on O'Reilly's Production, Editorial, and Tools teams. Currently, he works as a Publishing Technologies Specialist, maintaining O'Reilly's XML-based toolchain for generating EPUB and Mobi formats of both frontlist and backlist titles. He also helps coordinate O'Reilly's digital distribution efforts to electronic sales channels, and is currently assisting in R&D efforts surrounding HTML5

#### Hire Sanders Kleinfeld

**For Inquiries Contact**  
[talent@oreilly.com](mailto:talent@oreilly.com)

**Press Inquiries**  
[press@oreilly.com](mailto:press@oreilly.com)

**Find Other Authors**

Consultants

#### Also find Sanders on:



#### Recent Twitter posts:

- [@nbousquet](#) Most EPUBs out now use XHTML 1.1, so EPUB is already pretty Web-compliant. Maybe HTML5 ebooks will be more attractive to pirates? [2 days ago](#)

#### Author Events

**Webcast: HTML5 for Publishers**  
2014-11-20 14:30:14