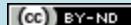


THE CONVERGENCE OF EPUB AND THE WEB

Ivan Herman, W3C

Buchtage Berlin 2015

2015-06-18



This work is licensed under a [Creative Commons Attribution 3.0 License](#), with attribution to [W3C](#) and [IDPF](#)

Copyright ©2015 W3C® (MIT, ERCIM, Keio, Beihang)

THESE SLIDES (IN HTML) ARE AVAILABLE ON THE WEB

See:

<http://w3c.github.io/dpub/buchtage-2015-2015/>

ABOUT W3C: “LEADING THE WEB TO ITS FULL POTENTIAL”

- Founded in 1994, and directed by Tim Berners-Lee, the Inventor of the Web
- Member organization with cca. 390 members
- Cca. 80 staff bound to “hosts” in the USA (MIT), China (Beihang University), Japan (Keio University), and France (ERCIM)
- Focused on Web ecosystem: users, developers, browsers, etc.
- W3C focuses both on the Open Web, as well as specific industry requirements brought by industry segments

Credit: Beehive Communications, 2007

DPUB IG ORIGINS

- The publishing industry is, probably, the most important user of W3C's Web technologies after (traditional) browsers:
 - almost all journals, magazines, etc, have an online version these days
 - scholarly publishing cannot exist without the Web any more
 - EPUB is, essentially, a frozen and packaged Web site
- The quality requirements of the publishing industry are very high:
 - high quality typesetting, graphics, etc.
 - new forms of publishing will be based on high level of interactions, rich media, ...
 - common document and data publishing comes to the fore

DPUB IG ORIGINS (CONT.)

- But... the publishing industry had been in an entirely “passive” mode v.a.v. Web technologies
 - no participation in the development of fundamental Web technologies
 - W3C (and other standard bodies) hardly know about the requirements that this industry may have
 - the potential synergy between Web developers and publishers is missed out
- Consequence: Working Groups at W3C set their priorities without knowing about, and considering, the publishing industry

DPUB IG ORIGINS (CONT.)

- W3C and IDPF organized a series of exploratory workshops in 2012 to create a missing synergy among communities
- The W3C [Digital Publishing Activity](#) and the [Digital Publishing Interest Group](#) was formally chartered in May 2013
- DPUB IG has weekly teleconferences and bi-annual face to face meetings

DPUB IG MISSION

- Experts familiar with the ins and outs of digital publishing and its associated industry groups identify issues that are not addressed by the Open Web Platform
- Our goal is to raise issues to W3C working groups who can update or develop specs based on the needs of the publishing community.
- DPUB IG has task forces with leads to focus attention on specific issues and collaborate with other WGs. We often delegate members to other WGs as well

See our [website](#) for more detail.

CURRENT DPUB IG TASK FORCES

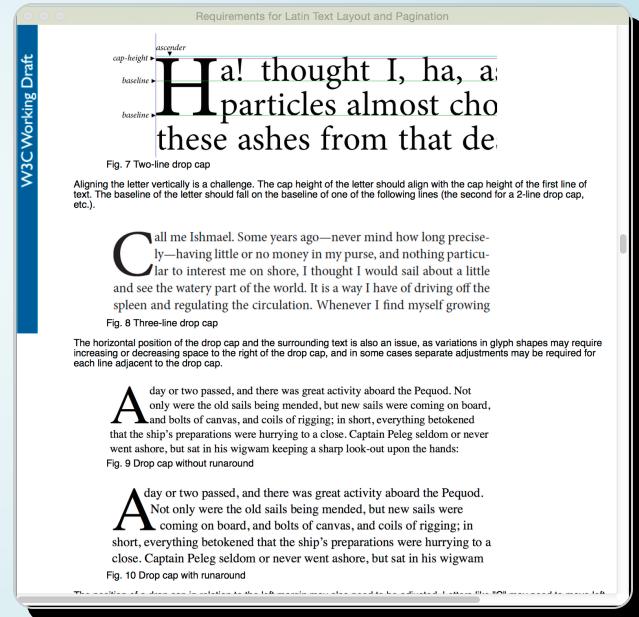
- *Layout and Styling (a.k.a Pagination)*: led by Dave Cramer
- *Identification*: led by Bill Kasdorf
- *Content and Markup (a.k.a. Structural Semantics)*: led by Tzviya Siegman
- *Accessibility*: led by Deborah Kaplan and Charles LaPierre
- *STEM*: led by Peter Krautzberger

PAST (CLOSED) DPUB IG TASK FORCES

- *Annotations*: led by Rob Sanderson
- *Metadata*: led by Bill Kasdorf and Madi Solomon

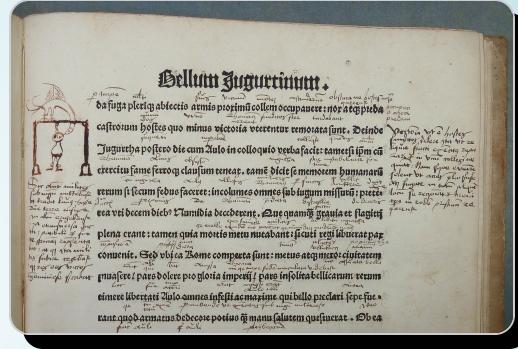
EXAMPLE: LAYOUT, STYLING, PAGINATION

- An evolving document:
“Requirements for Latin Text Layout and Pagination”
- Describes issues like hyphenation, spreads and bleeds, drop caps, pagination, etc.
- Has greatly influenced some current CSS Work, e.g. “[CSS Inline Layout Module Level 3](#)” (handling initial letters, dropcaps), or “[CSS Generated Content for Paged Media Module](#)” (handling running heads and footers)



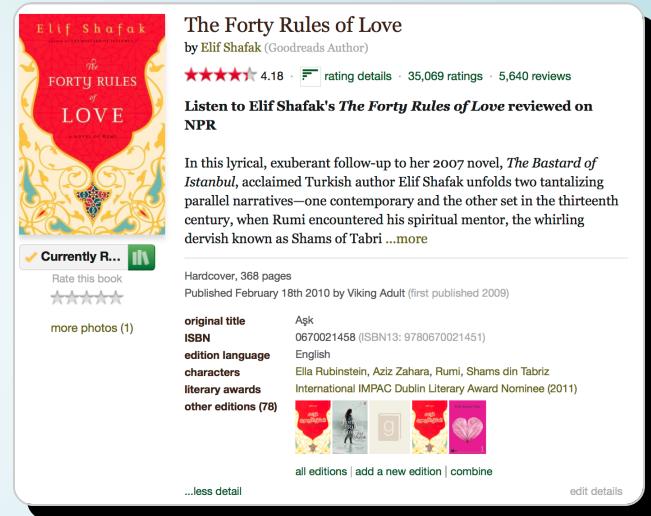
EXAMPLE: ANNOTATIONS

- Published an Annotation Use Cases
- Activity and work has shifted to the [Web Annotations Working Group](#)
 - the work aims at annotation for all forms of Web Documents, whether in a browser or an eBook
 - the first Working Group at W3C joining forces of both browser and publishing industries from the start



METADATA

- Interviewed publishing experts in metadata
- Published DPUB IG Metadata Task Force Report
 - a note summarizing findings in interviews
 - discovered that OWP has the tools that publishers seek, but many in publishing are not familiar with tools
 - this group can aid the publishing industry in using OWP tools to implement existing vocabularies



LOT HAS BEEN DONE

Fig. 10 Two-line drop cap

Aligning the letter vertically is a challenge. The cap height of the letter should align with the cap height of the first line of text. The baseline of the letter should fall on the baseline of one of the following lines (the second for a 2-line drop cap, etc.).

Annotation Use Cases

W3C Interest Group Note 17 November 2014

This version: <http://www.w3.org/TR/2014/NOTE-dpub-annotation-uc-20141117/>
Latest published version: <http://www.w3.org/TR/dpub-annotation-uc/>
Latest editor's draft: <http://w3c.github.io/dpub-annotation/>
Editor: Robert Sanderson, Stanford University, azaroth42@gmail.com
Repository: [Github Repository](#)

This document is also available in this non-normative format: [ePub](#)

Copyright © 2014 W3C® (MIT, ERCIM, Keio, Beihang). All Rights Reserved. W3C liability, trademark and document use rules apply.

Abstract

This document describes the set of use cases generated for Annotation and Social Reading within the W3C Digital Publishing Interest Group, in coordination with the Open Annotation Community Group.

WHAT'S NEXT?

- Continue our task forces with targeted deadlines and deliverables
- *Involve DPUB IG members in the planning and development of EPUB+WEB*
 - this may lead to the creation of new task forces

IS IT A BOOK? IS IT A WEB SITE?

3.3 Multiple Alternatives

Multiple if statements can be combined to evaluate complex decisions.

In Section 3.1, you saw how to program a two-way branch with an if statement. In many situations, there are more than two cases. In this section, you will see how to implement a decision with multiple alternatives.

For example, consider a program that displays the effect of an earthquake, as measured by the Richter scale (see Table 3).

Table 3 Richter Scale

Value	Effect
8	Most structures fall
7	Many buildings destroyed
6	Many buildings considerably damaged, some collapse
4.5	Damage to poorly constructed buildings

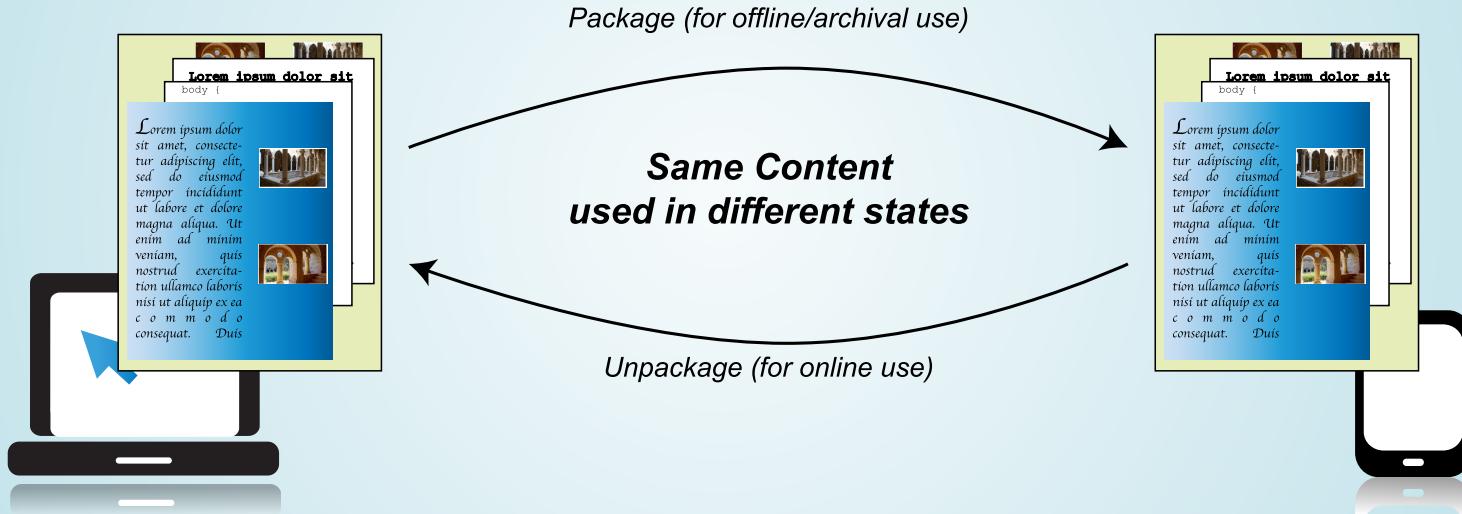
The Richter scale is a measurement of the strength of an earthquake. Every step in the scale, for example from 6.0 to 7.0, signifies a tenfold increase in the strength of the quake.



WHAT'S NEXT? EPUB+WEB

- IDPF and W3C published joint vision of future of digital publishing in a white paper entitled “*Advancing Portable Documents for the Open Web Platform: EPUB+WEB*”
 - White Paper is available at <http://w3c.github.io/epubweb/>
 - provide feedback at <https://github.com/w3c/epubweb/issues>

EPUB+WEB IS A VISION FOR THE FUTURE



THE VISION

- Portable documents are fully native citizen of the Web
- Separation between online (i.e., the “Web”) and portable (i.e., “EPUB”) is diminished to zero
- This means:
 - content authored for primarily offline use can be used online by loading it into a browser
 - content authored for primarily online use can be easily saved as a portable document for offline use
 - these should be doable smoothly, solely based on the user’s interaction

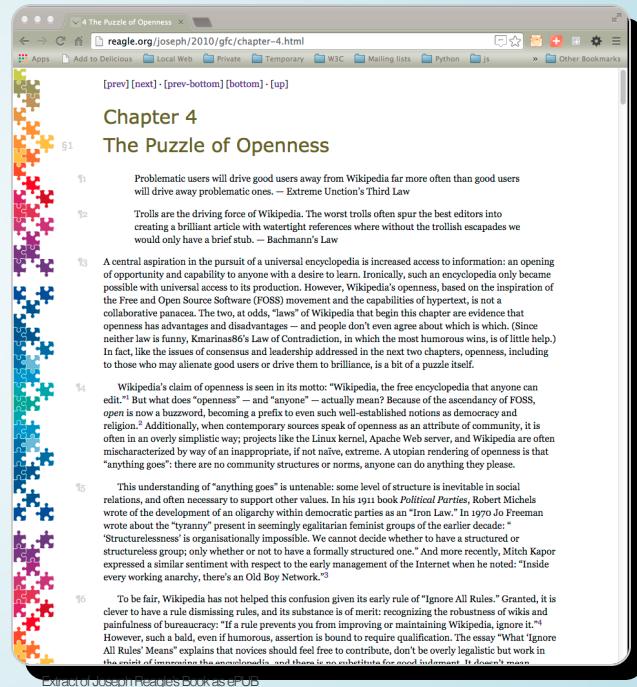
THE VISION

- Publishers can choose to utilize either or both of these publishing modes
- Users can choose either or both of these consumption modes
- Essential features flow seamlessly between on-line and off-line modes, like
 - cross-references, user annotations, access to on-line databases
 - licensing and rights management
 - etc.

WHY BOTHER?

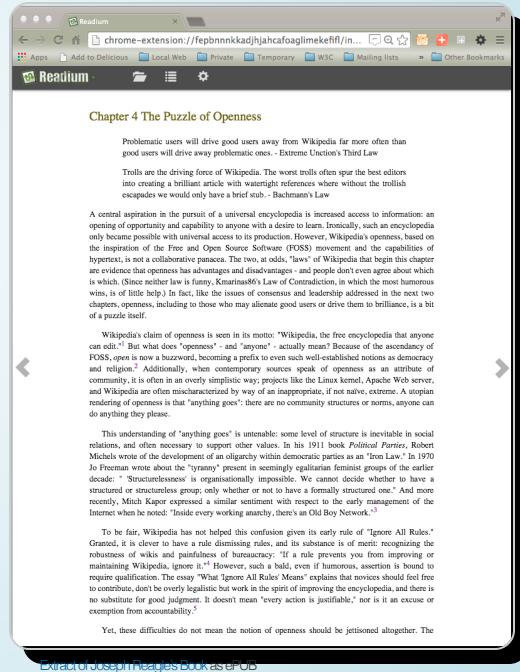
FOR EXAMPLE: BOOK IN A BROWSER

- On a desktop I may want to read a book just like a Web page:
 - easily follow a link “out” of the book
 - create bookmarks “into” a page in a book
 - use useful plugins and tools that my browser may have
 - create annotations



FOR EXAMPLE: BOOK IN A BROWSER (CONT.)

- But:
 - sometimes I may need the computing power of my desk-top for, e.g., interactive 3D content
 - at other times I may also want to use a small dedicated reader device to read the book on the beach...
- All these on *the same* book (not conversions from one format to the other)!



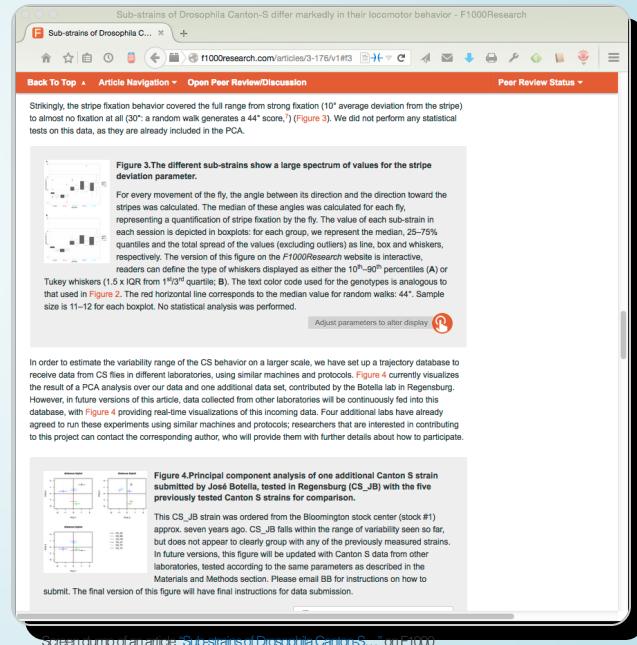
FOR EXAMPLE: I MAY NOT BE ONLINE...

- I may find an article on the Web that I want to review, annotate, etc., while commuting home on a train
- I want the results of the annotations to be back online, when I am back on the Internet
 - Note: some browsers have an “archiving” possibility, but they are not interoperable
 - the content can definitely not be read on a dedicated reader

Photo credit: Bryan Ong, flickr

FOR EXAMPLE: SCHOLARLY PUBLISHING

- My paper is published, primarily, online, but people may want to download it for offline use
- The format of the paper should be adaptable to my reading environment
 - do not want a two column, fixed layout file that I cannot handle on my iPad...
- My “paper” may also contain video, audio, data, programs...
 - scholarly publishing is not text only any more!



FOR EXAMPLE: IN-HOUSE PUBLISHING

- Major companies (IBM, Intel, Boeing, FAO, Renault,...) are specialized publishers through the publication of huge amount of documentation
- Delivering it on paper is not an option any more
- Fast refresh time is needed
- The same document should be available offline (e.g., in the cockpit) or online (e.g., on the work floor): there should be no difference between the two



Photo credit: Pedro Aguiar, Flickr

FOR EXAMPLE: EDUCATIONAL MATERIALS

- What is an educational publication?
 - a book of possibly long texts that requires offline access on dedicated devices?
 - a packaged application with built-in interactive tests, animated examples?
 - a Web client reaching out to Web services for assessing test results, to encyclopedia, ...?
 - an interactive data container storing various data for, e.g., demonstrations?
- The borderline between a “book” and a “(Web) Application” are becoming blurred!

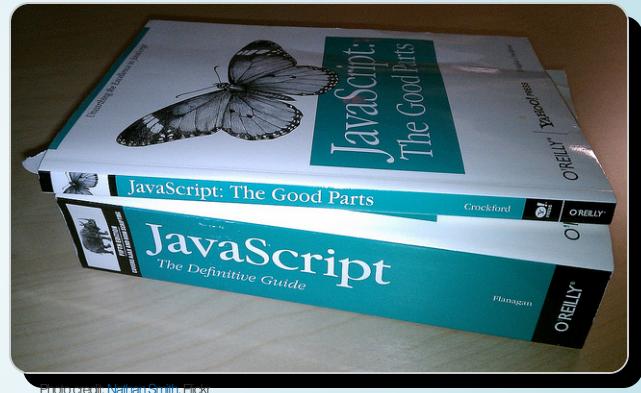


Photo credit: Merrill College of Journalism, Flickr

SYNERGY EFFECTS OF CONVERGENCE

ADVANTAGE FOR PUBLISHERS' COMMUNITY

- Publishers want to concentrate on what they know better: how to produce, edit, curate, etc, great content
- Publishers are not technology companies, nor do they intend to be; they want instead to rely on the vibrant Web community!

Photo credit: Nathan Smith, flickr

ADVANTAGE FOR PUBLISHERS' COMMUNITY (CONT.)

- OWP is more than “just” HTML, CSS, MathML, etc.
- It defines, e.g., a large number of facilities that provide access to, e.g., system resources or utilities
 - index database, Web storage, battery status API, real-time communication, geolocation,...
- Aligning on OWP means that publishing oriented devices, software, services, etc, can rely on those
 - instead of possibly re-inventing the wheel...

ADVANTAGE FOR THE WEB COMMUNITY

- Publishers have a long experience in ergonomics, typography, paging, complex layout, etc.
- Publishing long texts, with the right aesthetics, readability, structure, etc., is an expertise the Web community can profit from
- Experience of publishers in the complete workflow for producing content may become important for Web design

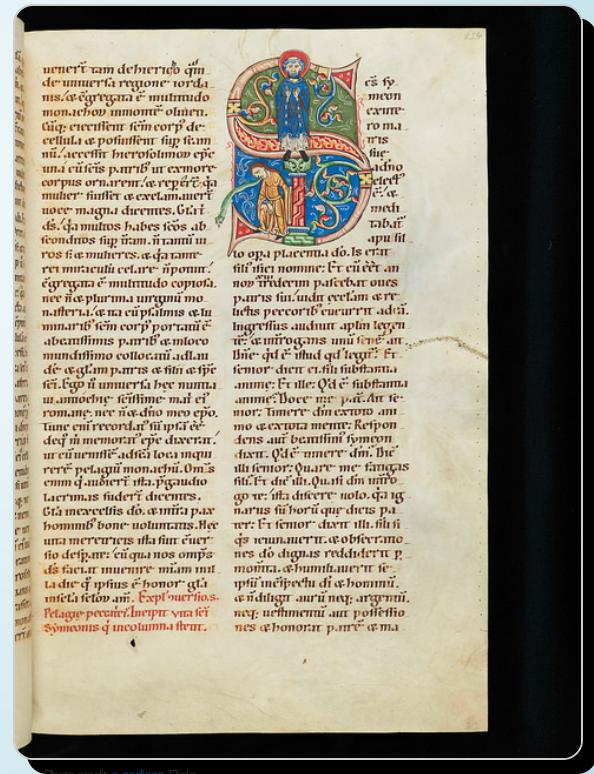


Photo credit: [Eduardo Sánchez](#)

HOW DO WE GET
THERE?
(TECHNICALLY)

HOW DO WE GET TO EPUB+WEB?

- A strong cooperation between the different communities should be ensured
- Technical challenges must be identified
 - note that some of the challenges are not EPUB+WEB specific, but EPUB in general (e.g., pagination control)
- Some examples follow...

TECHNICAL CHALLENGE: ARCHIVAL FORMAT

ROUGH STRUCTURE OF AN EPUB3 FILE

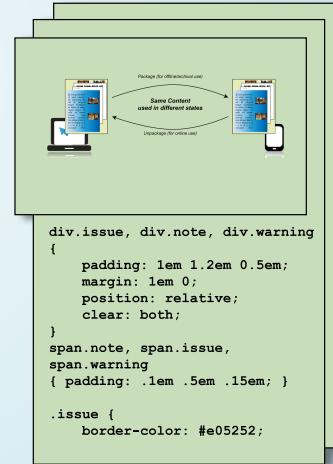
Administrative files
(manifest, metadata, rights,...)

```
<?xml version='1.0' encoding='utf-8'?>
<package
  xmlns="http://www.idpf.org/2007/opf"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  prefix="cc: http://creativecommons.org/ns#" unique-identifier="pub-id" version="3.0" xml:lang="en">
  <metadata>
    <dc:title>EPUB-WEB</dc:title>
    <dc:creator id="creator">Markus Gylling, Tzviya Siegman, Ivan Herman</dc:creator>
    <dc:language>en-US</dc:language>
    <rel="cc:license" />
  </metadata>
  <manifest>
    <item href="nav.xhtml" id="nav">
```

Content
(HTML, SVG, MathML,...)

```
<section>
  <h2>1. Our Vision</h2>
  <p>Our vision for EPUB-WEB is that portable documents become fully native citizens of the Open Web Platform. In this vision, the current format- and workflow-level separation between offline/portable (EPUB) and online (Web) document publishing is diminished to zero. These are merely two dynamic manifestations of the same publication: content authored with online use as the primary mode can easily be saved by the user for offline reading in portable document form. Content authored primarily for use as a portable document can be put online, without any need for refactoring the content. Publishers can choose to utilize either or both of these publishing modes,
```

Auxilliary files
(PNG, CSS, JS,...)



Open Packaging Format (OPF)

ARCHIVAL FORMAT

- There is a growing interest at W3C for a Web friendly packaging format:
 - should be streamable
 - should rely, as much as possible, on existing Web technologies (e.g., HTTP)
- Use cases include:
 - retrieve an HTML file with related CSS files, images
 - access Web Applications (“Widgets”) with all libraries involved
 - *EPUB+WEB is a clear use case*
- But: current Web Packaging proposal is *not* OPF based
- This may lead to a different packaging to EPUB+WEB

POSSIBLE EPUB+WEB PACKAGING STRUCTURE

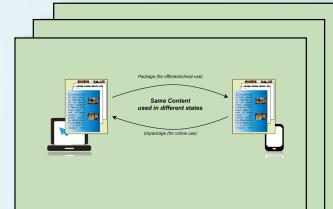
Administrative files
(manifest, metadata, rights,...)

```
{
  "metadata" : {
    "dc:title" : "EPUB+WEB",
    "dc:creator" : [
      "Markus Gylling",
      "Tzviya Siegman",
      "Ivan Herman"
    ],
    "dc:language" : "en-US"
  },
  "manifest" : ...
}
```

Content
(HTML, SVG, MathML,...)

```
<section>
  <h2>1. Our Vision</h2>
  <p>Our vision for EPUB+WEB is that portable documents become fully native citizens of the Open Web Platform. In this vision, the current format- and workflow-level separation between offline/portable (EPUB) and online (Web) document publishing is diminished to zero. These are merely two dynamic manifestations of the same publication: content authored with online use as the primary mode can easily be saved by the user for offline reading in portable document form. Content authored primarily for use as a portable document can be put online, without any need for refactoring the content. Publishers can choose to utilize either or both of these publishing modes,
```

Auxilliary files
(PNG, CSS, JS,...)



```
div.issue, div.note, div.warning {
  padding: 1em 1.2em 0.5em;
  margin: 1em 0;
  position: relative;
  clear: both;
}
span.note, span.issue,
span.warning {
  padding: .1em .5em .15em; }

.issue {
  border-color: #e05252;
```

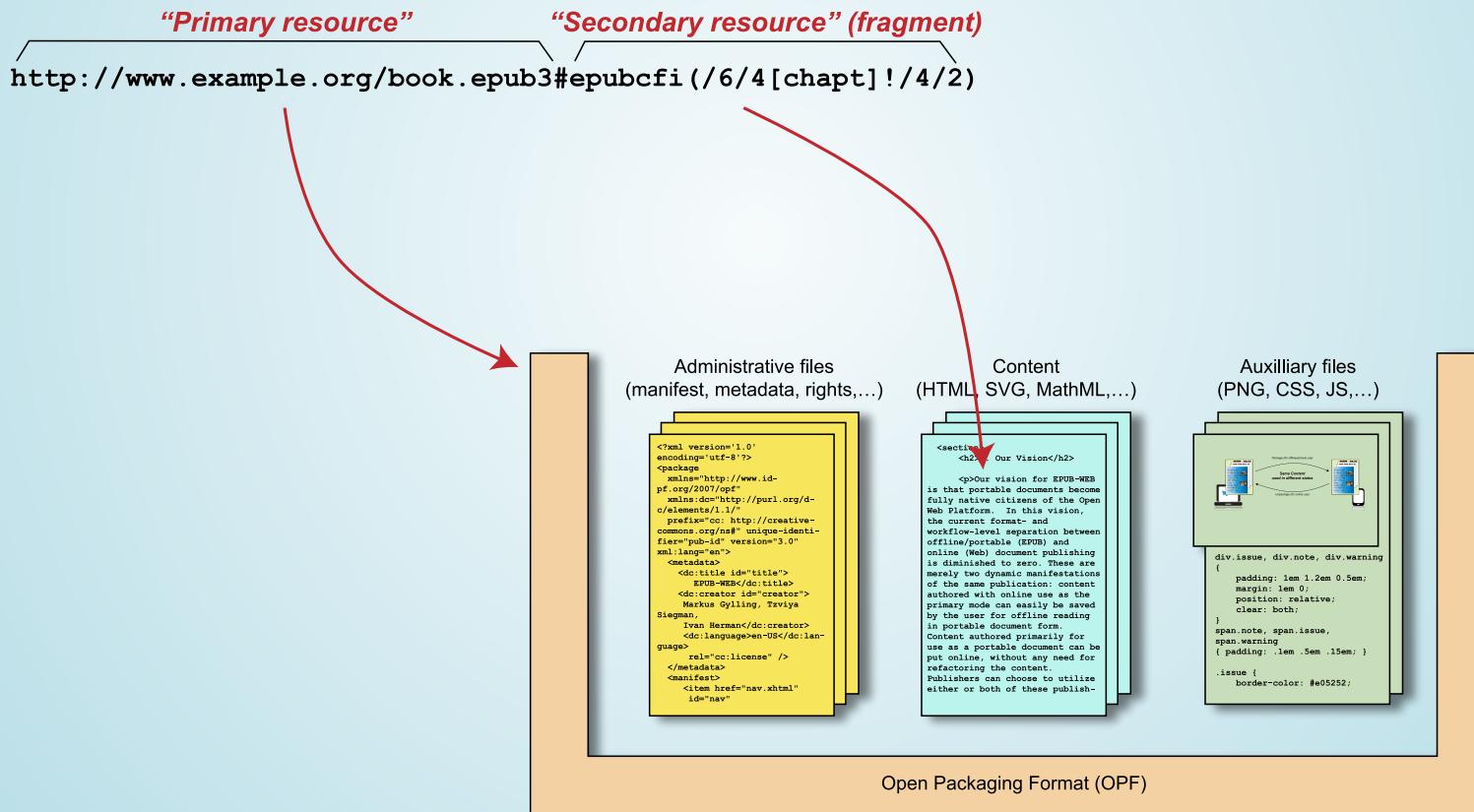
Web Packaging Format

WINDOW OF OPPORTUNITY IS OPEN NOW!

- The Web community does not concentrate strongly on the Web Packaging format yet
 - some other technologies, like Service Workers, must be clarified first
- But it may become a hot topic when some other technical challenges are solved
- The publishing community should get on board now to ensure that a format agreeable for all will be defined

TECHNICAL CHALLENGE: IDENTIFICATION

IDENTIFICATION: EPUB3 CASE



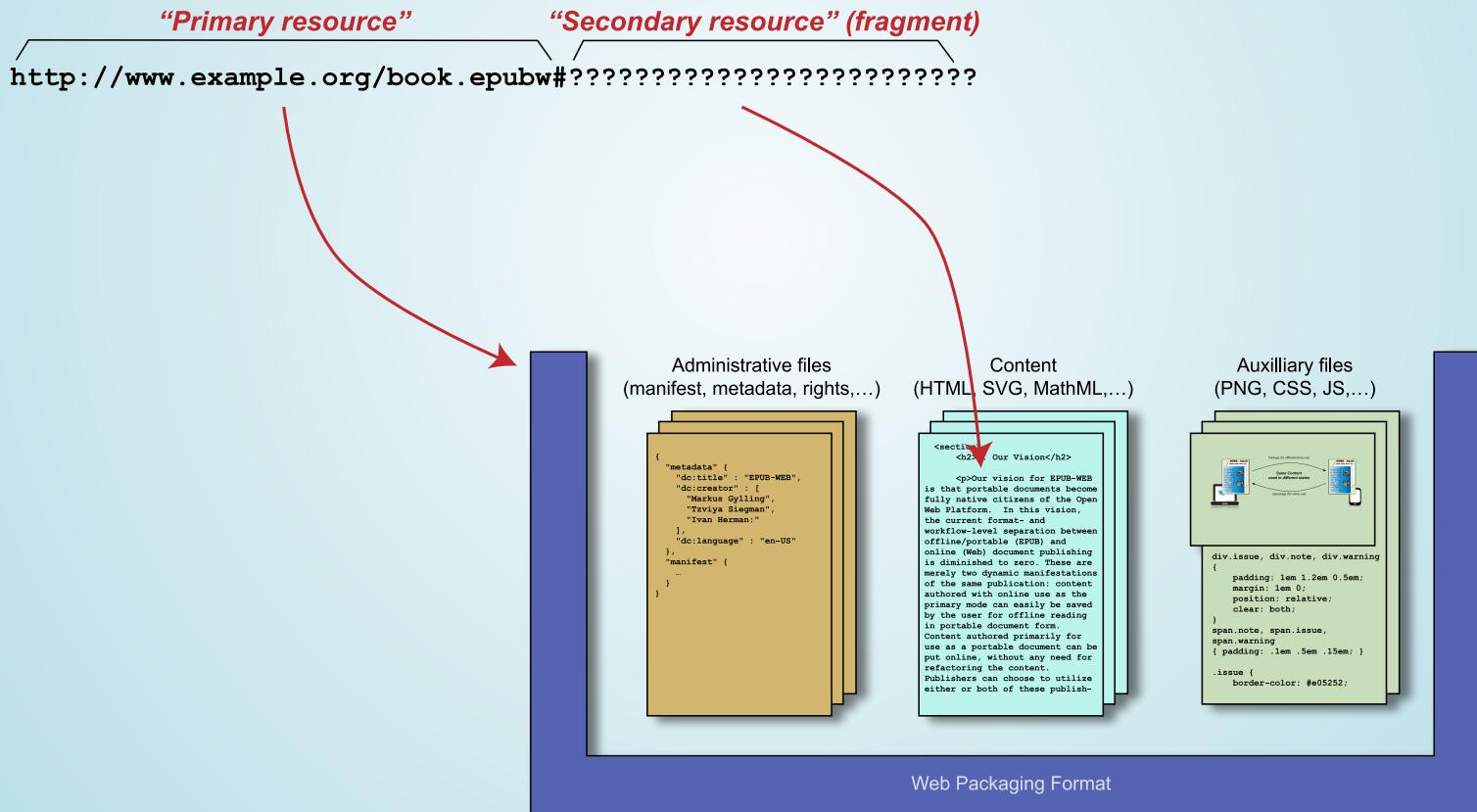
TWO PARTS OF IDENTIFICATION

1. Primary resource:

- This is a major discussion among publishers, libraries, archival organizations
- i.e., what is the URI for...
 - Shakespeare's Hamlet?
 - its digital edition published by Publisher XYZ?
 - the copy I own and annotate?
- It is not the role of IDPF or W3C to define those URL-s
 - though EPUB+WEB would push for much more “webby” URL schemes, i.e., HTTP(S) URLs

2. Secondary resource: IDPF has adopted the CFI standard for EPUB3

IDENTIFICATION: THE EPUB+WEB CASE



TECHNICAL CHALLENGE: PRESENTATION CONTROL

PRESENTATION CONTROL

- What is the level of user control of the presentation?
- The Web and eBook traditions are vastly different:
 - in a browser, the Web designer is in full control
 - CSS alternate style sheets are hardly in use
 - some user interface aspects can be controlled but only for the browser as a whole
 - in an eBook reader, there is more user control
 - foreground/background color
 - choice of fonts
- There is a need to reconcile these traditions

HOW DO WE GET
THERE?
(PRACTICALLY)

DPUB IG AND EPUB+WEB

- DPUB IG already has task forces dedicated to working on many of the EPUB+WEB issues
- The group will contribute to the formulation of the EPUB+WEB technical challenges, to a better understanding of the requirements
 - this may mean new task forces, e.g., for identification or packaging
- EPUB+WEB will become a guiding principle for the group's further work
- A [new charter](#) for the group is under preparation

IDPF, W3C, AND OTHERS

- On long term, some EPUB+WEB specification work has to be done
 - this requires a consensus and agreement of different communities
- IDPF and W3C (and maybe others?) may create the necessary groups, eventually

HOWEVER...

- EPUB+WEB does *not* replace EPUB3 (and upcoming EPUB3.1) at this moment
- Many of the new features will also be part of EPUB3.1 (e.g., structural semantics)
- The vision is a convergence of the EPUB3.* specifications and EPUB+WEB, eventually

CONCLUSION

- There is a great potential in a convergence between the Open Web Platform and Portable Documents
- It will require a common effort and cooperation of both communities
- But it is an exciting prospect!



SOME REFERENCES

DPUB IG Wiki

https://www.w3.org/dpub/IG/wiki/Main_Page

EPUB+WEB White paper:

<http://w3c.github.io/epubweb/>

EPUB+WEB Issue list:

<https://github.com/w3c/epubweb/issues>

New IG charter (in preparation) and the relevant issues' list:

<http://w3c.github.io/dpub-charter/index.html>

<https://github.com/w3c/dpub-charter/issues>

This presentation:

<http://w3c.github.io/dpub/buchtage-2015/index.html>

(PDF is also available for download)

DIRECT CONTACT

Ivan Herman, W3C

ivan@w3.org

<http://www.w3.org/People/Ivan/>

**THANK YOU FOR
YOUR ATTENTION!**