THE CONVERGENCE OF EPUB AND THE WEB

Tzviya Siegman, Wiley, Markus Gylling, IDPF, Ivan Herman, W3C

IDPF Digital Book Conference, New York, NY, USA

2015-05-28

This work is licensed under a Creative Commons Attribution 3.0 License, with attribution to W3C and IDPF

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

THESE SLIDES ARE AVAILABLE ON THE WEB

See:

http://w3c.github.io/dpub/idpf-digital-book-2015/

(Slides are in HTML)

REMINDER: W3C'S DIGITAL PUBLISHING INTEREST GROUP MISSION

> *The mission of the Digital Publishing Interest Group (DPUB IG) […] a forum for experts in the digital publishing ecosystem […] for technical discussions, gathering use cases and requirements to align the existing formats and technologies (e.g., for electronic books) with those used by the Open Web Platform […]*

DPUB IG MISSION EXPLAINED

- 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.

DIVISION OF ROLES

IDPF	W3C
Standards for the Electronic Publishing and Content Consumption (EPUB)	Standards for the General Web Technologies
Builds on foundational Web (e.g., W3C) Standards	Builds on foundational Internet (e.g., IETF, ECMA) Standards
Does *not* develop standards beyond publishing	Does *not* develop industry specific standards if there is another home for those

The key is *collaboration*.

CURRENT DPUB IG TASK FORCES

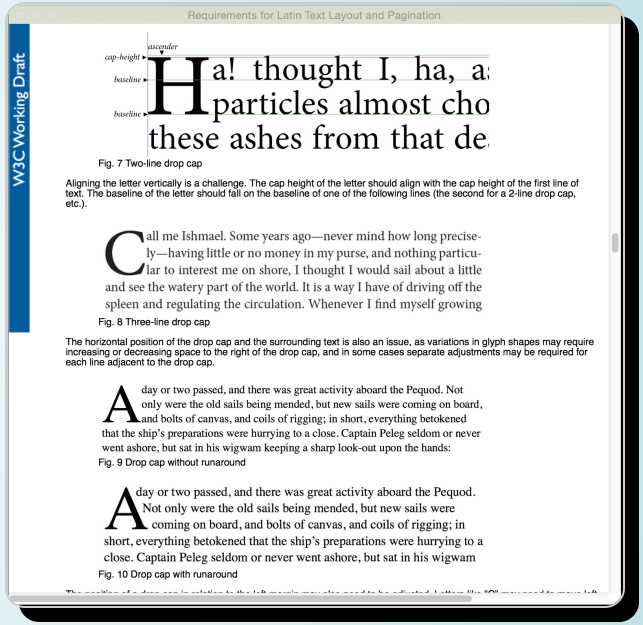
- *Layout and Styling (a.k.a Pagination):* led by Dave Cramer
- *Identification:* led by Bill Kasdorf
- *Content and Markup (a.k.a. Structural Semantics):* led by Tzviya Siegman
- *Accessibility:* led by Deborah Kaplan and Charles LaPierre
- *STEM:* led by Peter Krautzberger

PAST (CLOSED) DPUB IG TASK FORCES

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

EXAMPLE: LAYOUT AND STYLING

- 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: LAYOUT AND STYLING: HOUDINI PROJECT

> *The CSS-TAG "Houdini Project" to […] explain existing CSS functionalities better and make development of CSS polyfills […] easier by defining low level APIs of styling and layout on the web*

Credit: Vivliostyle blog, Toru Kawakubo

- The results may be essential for a proper (and standard) access to page-related information
- Pagination is a use case that goes beyond books. Slides, cards, and other "boxes" use the same functionality.
- Representatives of the Digital Publishing Community have been involved from the start.

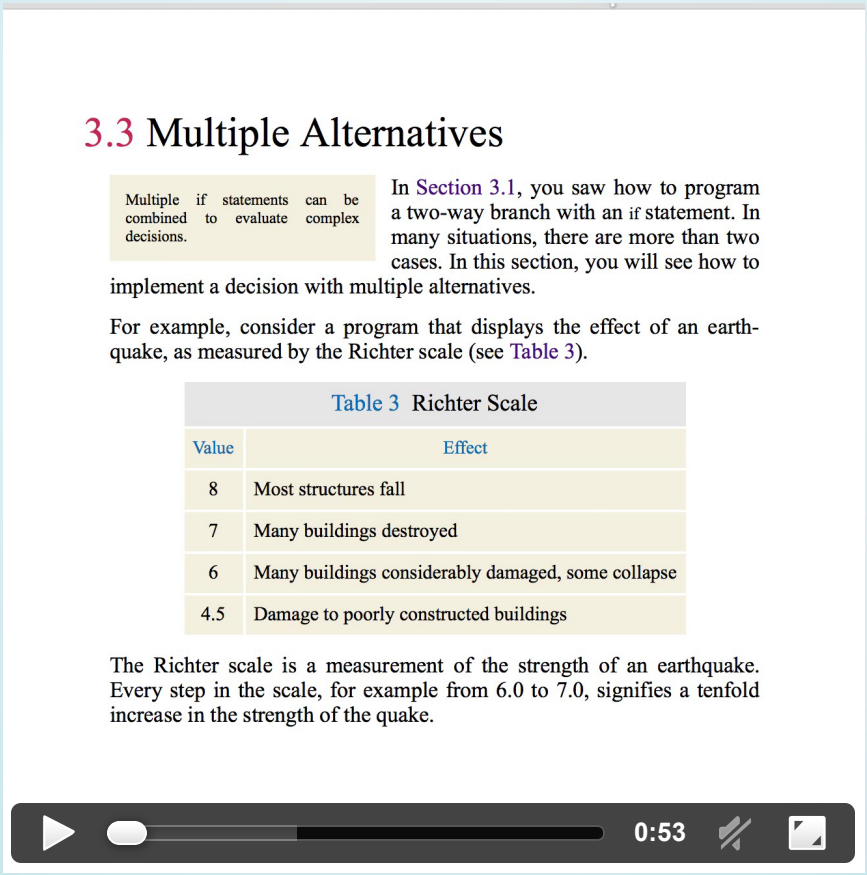
WHAT'S NEXT?

- 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

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`

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.

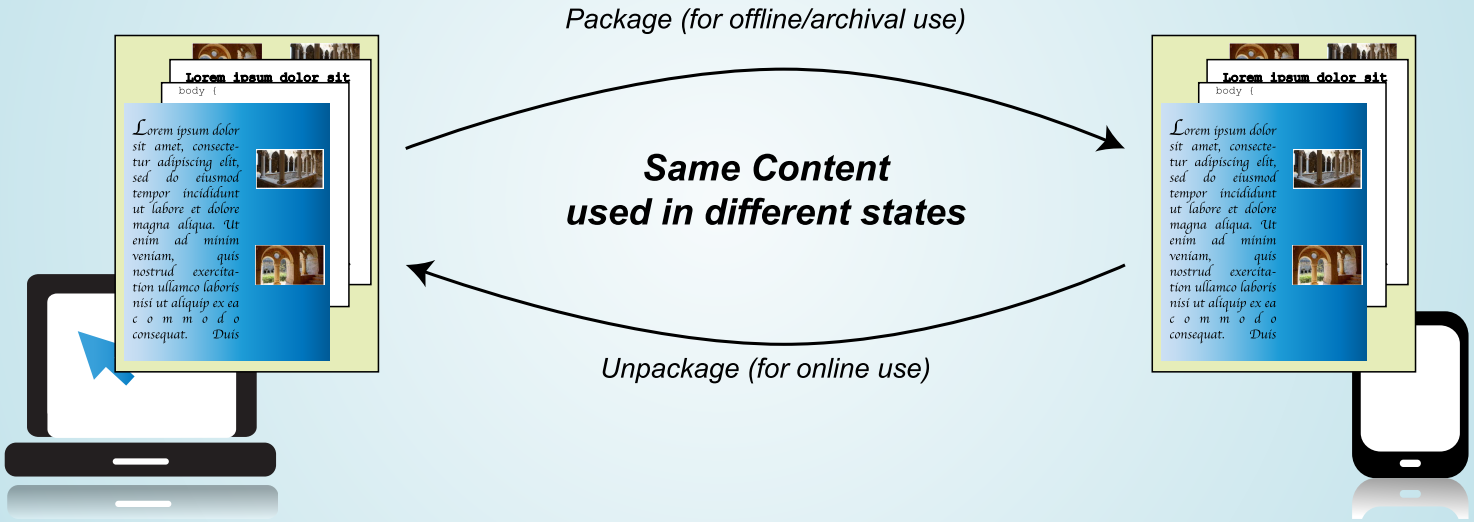
For example, consider a program that displays the effect of an earthquake, as measured by the Richter scale (see Table 3).

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.

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

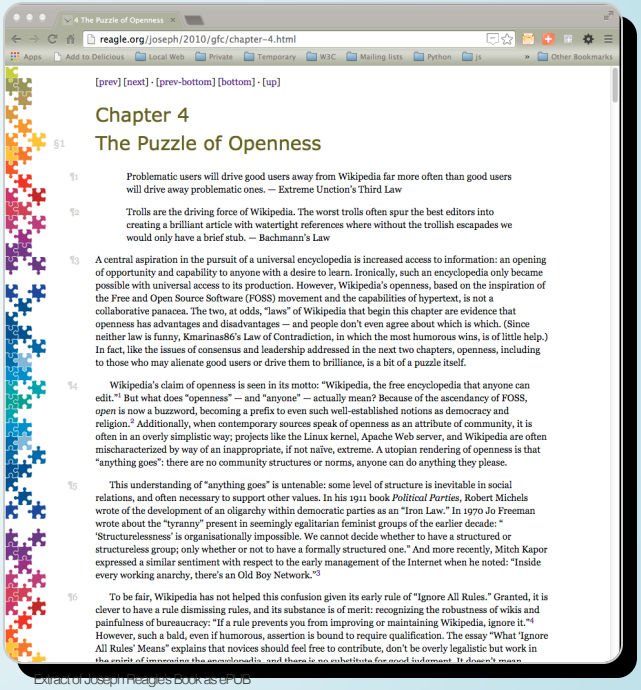
THE VISION

- 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

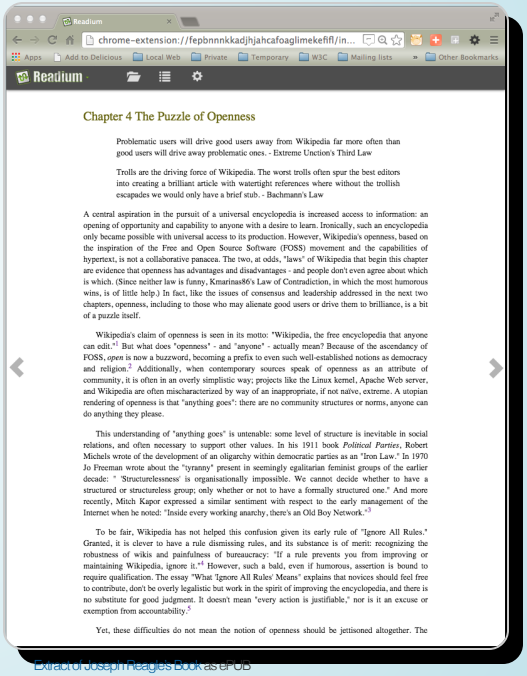
- On a desktop I may want to read a book just like a Web page:
  - easily follow a link “out” of the book
  - create bookmarks “into” a page in a book
  - use useful plugins and tools that my browser may have
  - create annotations



Extract of Joseph Reagle's book as ePUB

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)!



Extract of Joseph Reagle's Book as ePUB

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

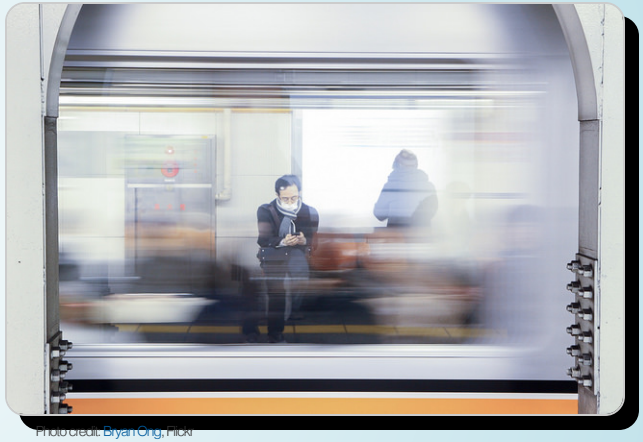
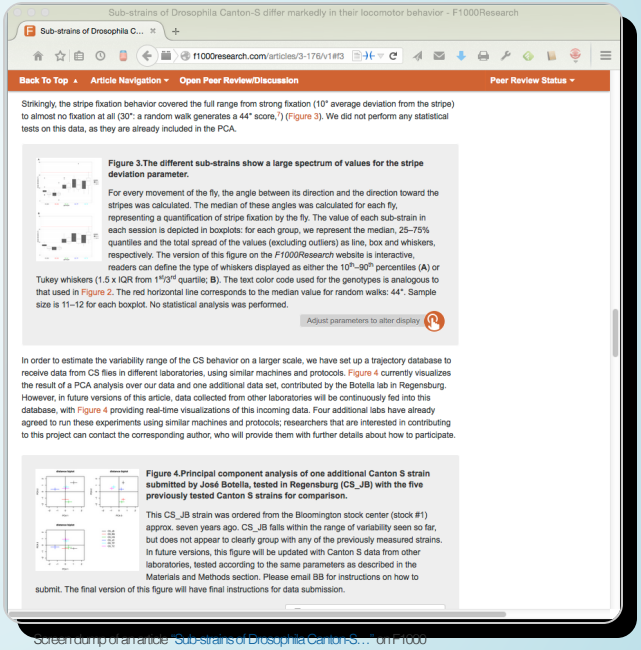


Photo credit: Bryan Ong, Flickr

FOR EXAMPLE: SCHOLARLY PUBLISHING

- My paper is published, primarily, on-line, 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: Pablo Agüera, 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: Memfi 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!

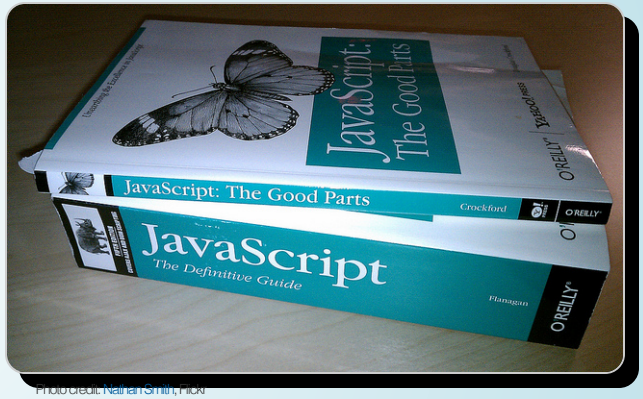


Photo credit: Nathan Shiff, Flickr

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

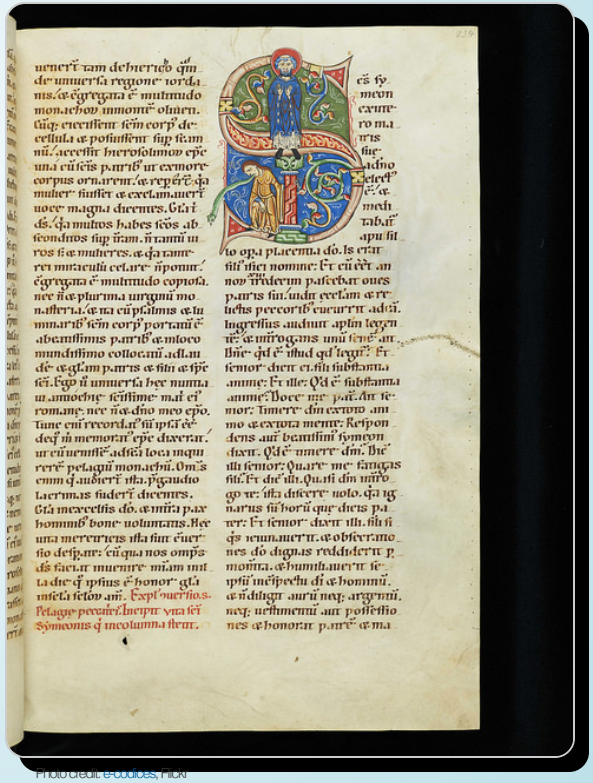


Photo credit: e-codices, Flickr

SOME COMMUNITIES THAT MAY BE AFFECTED

(Trade) Publishers
Browser vendors
Governmental bodies
Consumers of ebooks
Web designers
Web Developers

STM Publishers
Large companies
International institutions
Scholarly authors
Archivists
Publishing workflows

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' encod-
ing='utf-8'?>
<package
  xmlns="http://www.id-
pf.org/2007/opf"
  xmlns:dc="http://purl.org/d-
c/elements/1.1/"
  prefix="cc: http://creative-
commons.org/ns#" unique-identi-
fier="pub-id" version="3.0" xm-
l:lang="en">
  <metadata>
    <dc:title id="title">
       EPUB-WEB</dc:title>
    <dc:creator id="creator">
     Markus Gylling, Tzviya
Siegman,
     Ivan Herman</dc:creator>
    <dc:language>en-US</dc:lan-
guage>
       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 work-
flow-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. Con-
tent authored primarily for use
as a portable document can be put
online, without any need for re-
factoring 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;
```

Open Document Format (ODF)

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* ODF based
- This may lead to a different packaging to EPUB-WEB

EPUB-WEB PACKAGING STRUCTURE

Administrative files
(manifest, metadata, rights,…)

```
<?xml version='1.0' encod-
ing='utf-8'?>
<package
  xmlns="http://www.id-
pf.org/2007/opf"
  xmlns:dc="http://purl.org/d-
c/elements/1.1/"
  prefix="cc: http://creative-
commons.org/ns#" unique-identi-
fier="pub-id" version="3.0" xm-
l:lang="en">
  <metadata>
    <dc:title id="title">
      EPUB-WEB</dc:title>
    <dc:creator id="creator">
     Markus Gylling, Tzviya
Siegman,
     Ivan Herman</dc:creator>
     <dc:language>en-US</dc:lan-
guage>
      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 work-
flow-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. Con-
tent authored primarily for use
as a portable document can be put
online, without any need for re-
factoring 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

ACTUALLY…

- Moving to the Web would probably require to rethink the format of the administrative files
  - E.g., change formats from XML to JSON
- Note that this is already considered for EPUB3.1

EPUB-WEB PACKAGING STRUCTURE

Administrative files
(manifest, metadata, rights,…)

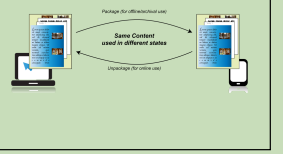
```
{
  "metadata" {
    "dc:title" : "EPUB-WEB",
    "dc:creator" : [
      "Markus Gylling",
      "Tzviya Siegman",
      "Ivan Herman:"
    ],
    "dc:language" : "en-US"
  },
  "manifest" {
    …
  }
}
```

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 work-
flow-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. Con-
tent authored primarily for use
as a portable document can be put
online, without any need for re-
factoring 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;
```

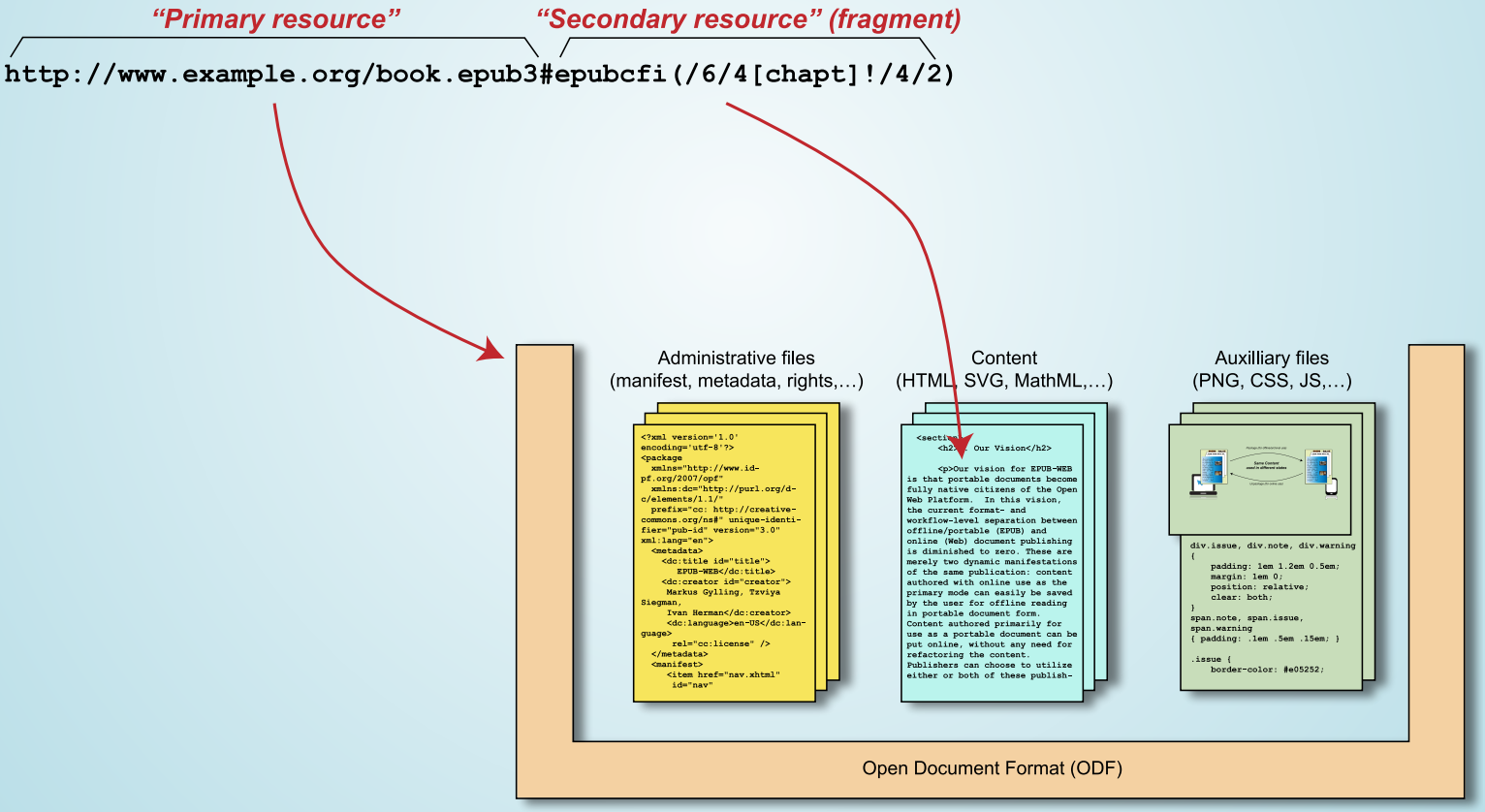
Web Packaging Format

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 challanges 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

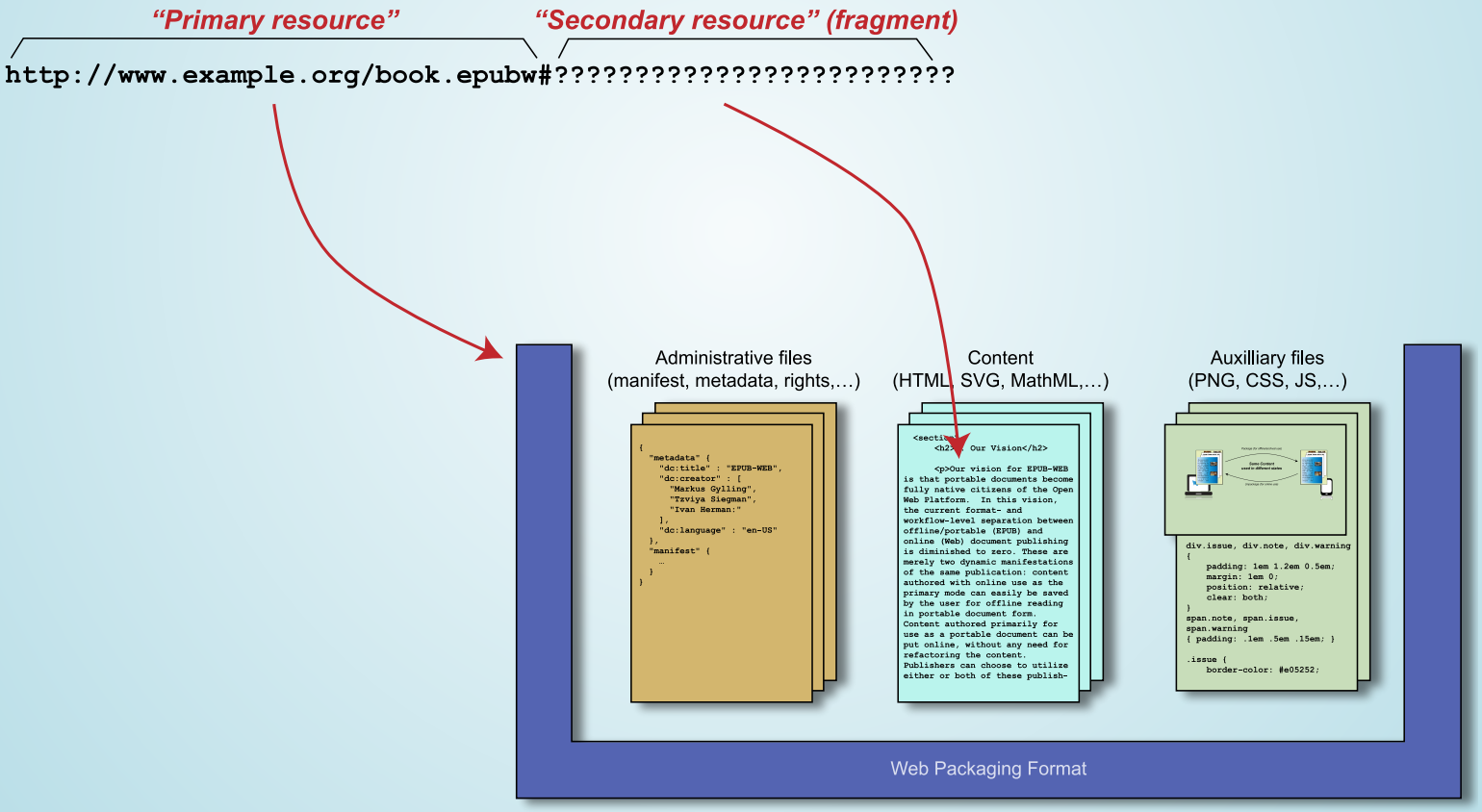
1. *Primary resource:*
   - This is a major discussion among publishers, libraries, archival organizations
   - i.e., what is the URI for…
     - Shakespeare's Hamlet?
     - its digital edition published by Publisher XYZ?
     - the copy I own and annotate?
   - It is *not* the role of IDPF or W3C to 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

POTENTIAL ISSUES WITH CFI

- CFI is deeply rooted in XML; however, EPUB-WEB should probably go for JSON and HTML5 for, e.g., package manifests
- CFI is bound to the current EPUB3 structures using manifest files
- CFI can be used with SVG content, but not with audio, video, data,…

An EPUB-WEB version for fragments should be prepared to those issues.

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

IDPF, W3C, AND OTHERS

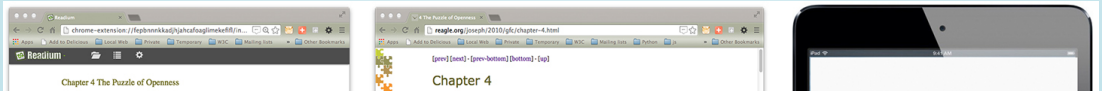
- On long term, some EPUB-WEB specification work has to be done
  - this requires a consensus and agreement of different communities
- IDPF and W3C (and maybe others?) may create the necessary groups, eventually

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

This presentation:

http://w3c.github.io/dpub/idpf-digital-book-2015/index.html (PDF is also available for download)

DIRECT CONTACT

Tzviya Siegman, Wiley
tsiegman@wiley.com
Markus Gylling, IDPF
mgylling@idpf.org
Ivan Herman, W3C
ivan@w3.org

THANK YOU FOR YOUR ATTENTION!