

# THE FUTURE OF W3C DIGITAL PUBLISHING

Ivan Herman, W3C

“Día del W3C en España 2015: el futuro de la edición digital”, Madrid,  
Spain

2015-10-14

# THESE SLIDES ARE AVAILABLE ON THE WEB

---

See:

<http://w3c.github.io/dpub/dia-del-w3c-2015-10/index.html/>

(Slides are in HTML)

# 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 this 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 many 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 Interest Group was formally created 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
- Goal is to raise issues to W3C working groups who can update or develop specs based on the needs of the publishing community.
- Work on a future vision of Digital Publishing called, tentatively, “Portable Web Publications (PWP)”

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

# IDPF AND W3C

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

The key is collaboration.

### 2.1.3 Methodology

One-to-one half-hour interviews were conducted over a four week period. These were casual and candid conversations on experiences or observations around the use of digital technologies in education.

### 2.2 Results

When the results of the W3C Metadata Task Force were combined with those from the interviews, a picture of what was common between Trade and Education publishing surfaced. The majority of those still in its transformative state to digital, were more focused on modularizing content.

#### Trade

Trade publishers stated that metadata complexity, mostly with ONIX, was a challenge to their business.

Trade publishers lamented the many metadata vocabularies (ONIX, PRISM, etc.) and the difficulty in keeping up-to-date on all of them.

"Few books are online anyway," was a general response from Trade publishers. Books were considered whole products and rarely modularised or componentized content. This was relegated to Title and Author and not much more.

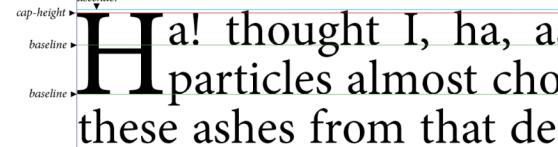


Fig. 10 Two-line drop cap

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.).

**C**all me Ishmael. Some years ago—never mind how long precisely—having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little

and see the watery part of the world. It is a way I have of driving off the

# SOME RESULTS OF THE PAST TWO YEARS

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](mailto: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

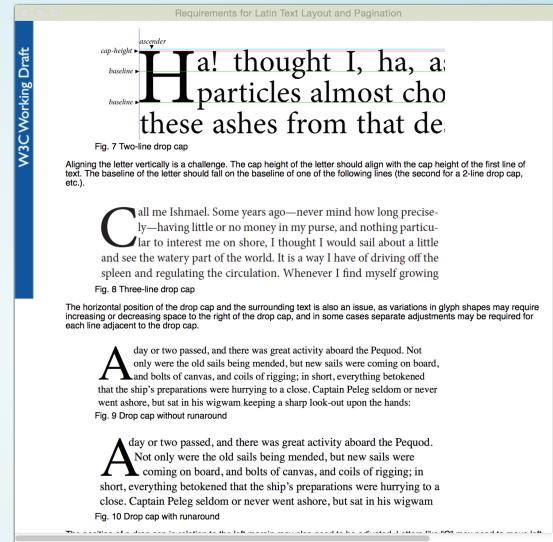
This document describes the set of use cases generated for Annotation and Social Reading within the W3C Digital Publishing Interest Group, in coordination with the Open Annotation Community Group.

#### foreword (role)

An introductory section that precedes the work, typically not written by the author of the work.

# 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)



# PRIORITIES FOR CSS...

- Another evolving document:  
“Priorities for CSS from the DPUB IG”
  - major emphasis on pagination-related features
- Provides a list of the top CSS priorities, and their current availability
- Also influences the work of the CSS Working Group

The screenshot shows a browser window displaying the W3C Working Draft page for "Priorities for CSS from the Digital Publishing Interest Group". The page is dated 20 August 2015. It includes sections for "2. List of Priorities" and "2.1. Features requiring implementation". A table lists various CSS features with their priority, specification status, and implementation status across different browsers.

Feature	Priority	Spec, Status	Implemented	Not Implemented	Polyfill?
font-feature-settings	5	[css3-fonts], CR	Firefox, IE, Chrome, Opera, Prince, AH	Safari (now implemented, but not in time for Q3 2015 releases)	No
hyphens	5	[css3-text], LC	Firefox, IE, Opera, Safari, Prince, AH	Chrome (bug filed; assigned on August 12, 2015)	Yes
vertical text	5	[css3-writing-modes], CR	Chrome, IE, Opera, Safari, AH	Firefox (which has announced intent to ship)	
Regions	5	[css-regions-1], WD	IE, Safari	Chrome, Firefox, Opera	Yes
Ruby	4	[css-ruby-1], WD	Firefox, partial support in Chrome, IE, Safari, Opera	unknown	Yes
Emphasis	4	[css3-text-decor], CR	Safari, partial support in Firefox, IE		

# CONTENT AND MARKUP

---

- Goal: identify the semantics of the elements in the HTML source beyond what the basic HTML tags can provide
  - “abstract”, “indexed term”, “footnote”, “chapter”, ...
- Express structural information (“where can that element be used”)
- Do it in a forward looking way in terms of W3C standards.
  - i.e., move away from `epub:type` used in EPUB 3
  - the resulting HTML should be valid
- These terms may be useful for the Web at large!

# CONTENT AND MARKUP: APPROACH CHOSEN

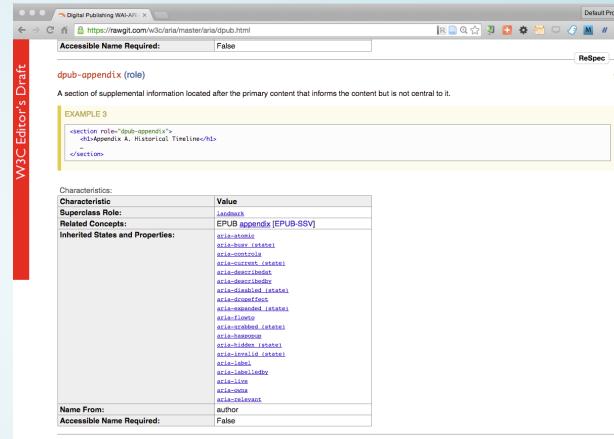
---

- Use “Accessible Rich Internet Applications (WAI-ARIA)” as a basic mechanism:
  - use specific attributes in HTML
  - attribute values convey a specific semantics
  - Assistive Technologies make use of the values

# CONTENT AND MARKUP: DPUB ARIA MODULE

- A Digital Publishing ARIA module is in development
- Publishing terms become part of ARIA

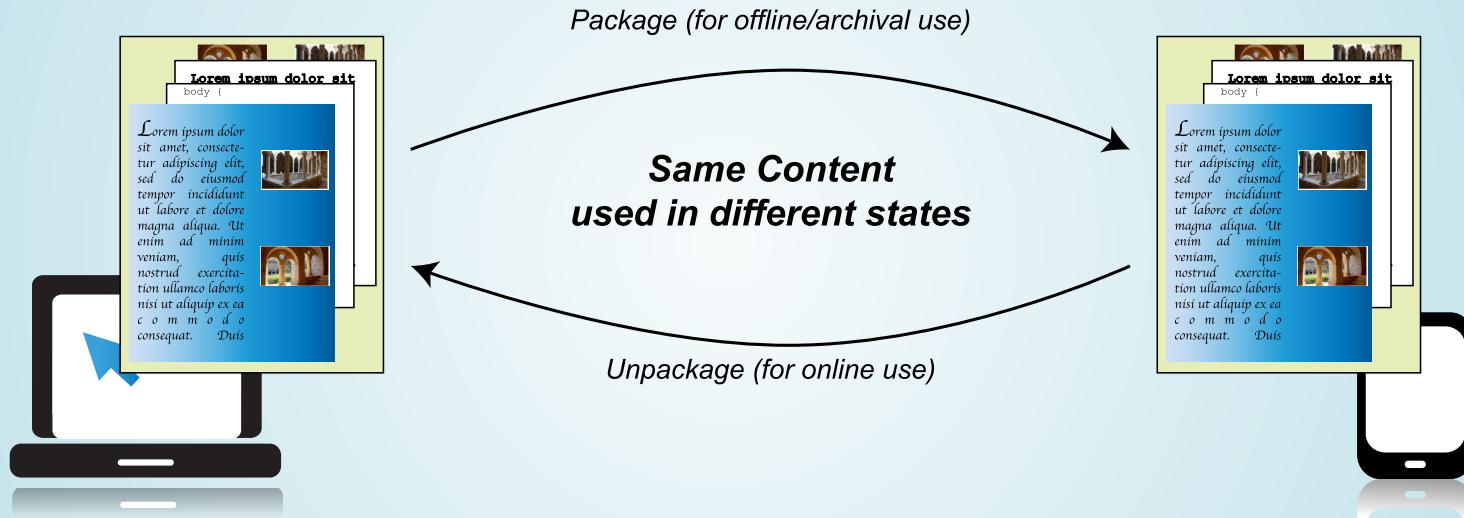
```
<section role="dpub-appendix">
  <h1>Appendix A. Historical Timeline</h1>
  ...
</section>
```



The background of the image features a large, rugged mountain peak, likely Glass House Mountains, with a distinctively rounded top. The mountain is covered in dense green vegetation at its base, transitioning to a rocky, reddish-brown surface higher up. In the foreground, there's a valley with a mix of green fields and dark, dense forests. The sky above is a uniform, pale blue-grey.

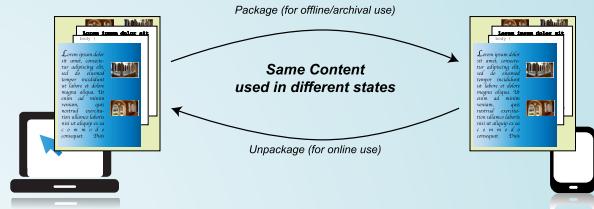
MAJOR WORK COMING  
UP: PORTABLE WEB  
PUBLICATIONS (PWP)

# PWP IS A VISION FOR THE FUTURE



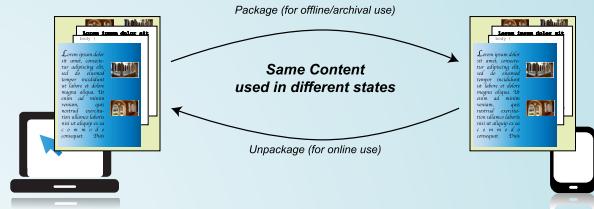
# THE VISION

- Portable Web Publication are fully native citizen of the Web
- Separation between online Web sites and offline/packaged (e.g., “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.

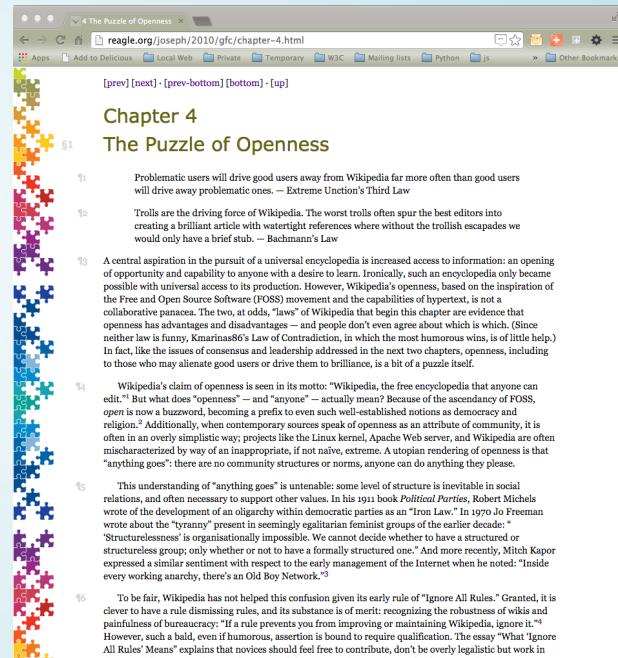


why ?



# FOR EXAMPLE: BOOK IN A BROWSER

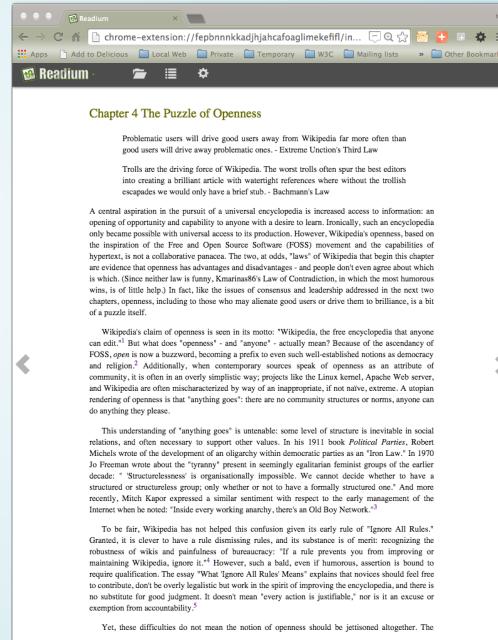
- On a desktop I may want to read a book just like a Web page:
  - easily follow a link “out” of the book
  - create bookmarks “into” a page in a book
  - use useful plugins and tools that my browser may have
  - create annotations



Credit: 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)!



Credit: Extract of Joseph Reagle's Book as ePUB

# FOR EXAMPLE: I MAY NOT BE ONLINE...

---

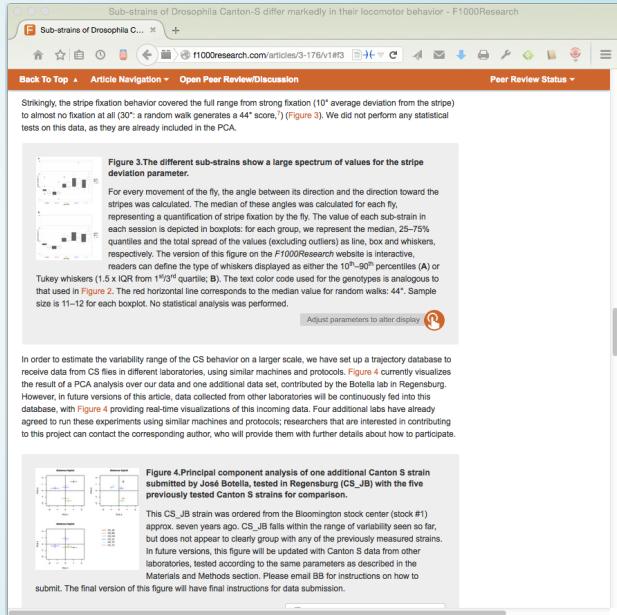
- I may find an article on the Web that I want to review, annotate, etc., while commuting home on a train
- I want the results of the annotations to be back online, when I am back on the Internet
  - Note: some browsers have an “archiving” possibility, but they are not interoperable
    - the content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



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!



Credit: Screen dump of an article "Sub-strains of Drosophila Canton-S..." on F1000

# FOR EXAMPLE: EDUCATIONAL MATERIALS

---

- What is an educational publication?
  - a book that requires offline access?
  - 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!



Credit: Merrill College of Journalism, Flickr

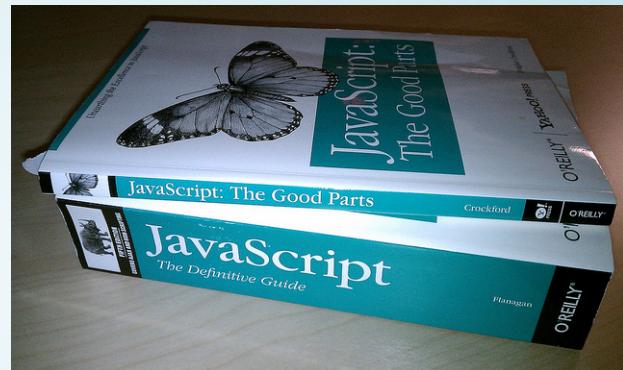
A close-up photograph of two hands shaking. One hand is light-skinned and wearing a white shirt cuff, while the other is dark-skinned. The hands are clasped firmly, symbolizing agreement or convergence.

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!



Credit: Nathan Smith, Flickr

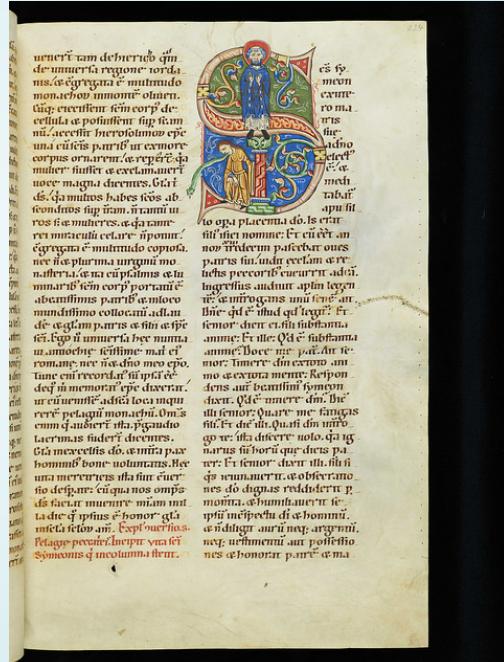
## ADVANTAGE FOR PUBLISHERS' COMMUNITY (CONT.)

---

- OWP is more than “just” HTML, CSS, MathML, etc.
- It also defines 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 more 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, ...
- 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



Credit: [e-codices](#), Flickr



HOW DO WE GET  
THERE?  
(TECHNICALLY)

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

**WARNING:  
EVERYTHING I SAY IS  
SUBJECT TO CHANGE!**



A photograph of the Guggenheim Museum Bilbao's titanium-clad facade. Overlaid on the image is a large, bold, yellow text block containing the title.

# TECHNICAL CHALLENGE: OVERALL ARCHITECTURE

# AN ASIDE: WEB AND SERVICE WORKERS

---

- Web Worker: a truly parallel thread within the browser
- A Service Worker is a special type of Web Worker, with additional features:
  - it is a *programmable network proxy*: the main thread's network calls are caught and the request/answer can be modified on-the-fly behind the scenes
  - it has an interface to handle a local cache for networked data

# AN ASIDE: WEB AND SERVICE WORKERS

---

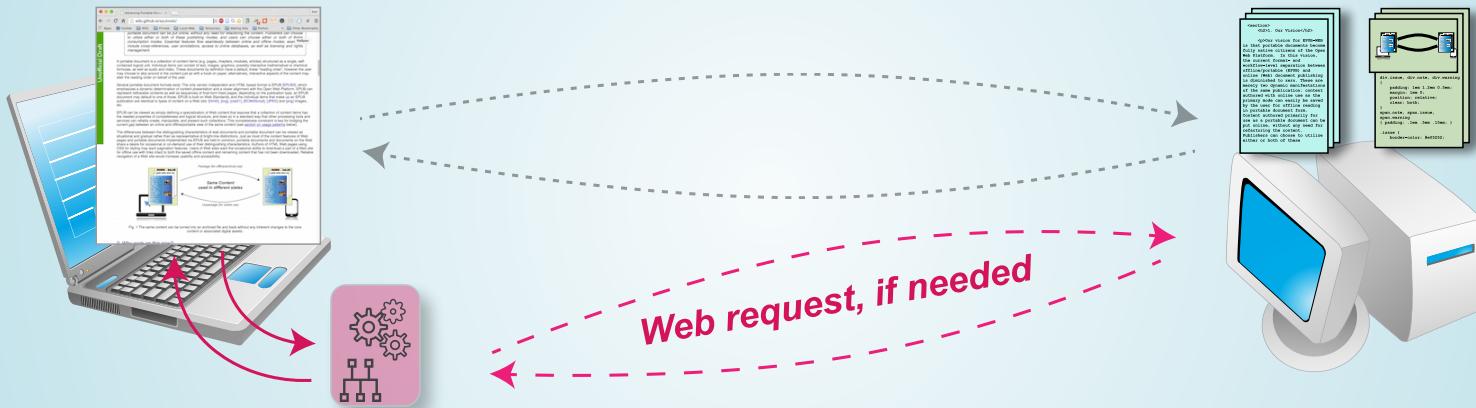
- Web Worker: a truly parallel thread within the browser
- A Service Worker is a special type of Web Worker, with additional features:
  - it is a *programmable network proxy*: the main thread's network calls are caught and the request/answer can be modified on-the-fly behind the scenes
  - it has an interface to handle a local cache for networked data

**Work in progress**

# ENVISIONED ARCHITECTURE: UNPACKED STATE



# ENVISIONED ARCHITECTURE: CACHED STATE



# ENVISIONED ARCHITECTURE: PACKED STATE



*Requests proxied through Service Worker  
content unpacked on-the-fly  
and possibly cached*

# ENVISIONED ARCHITECTURE: PACKED STATE



# TECHNICAL CHALLENGE: ARCHIVAL FORMAT



# ROUGH STRUCTURE OF AN EPUB3 FILE

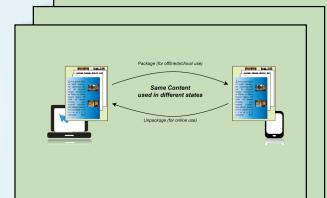
Administrative files  
(manifest, metadata, rights,...)

```
<?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 id="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,...)



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

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

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
  - *Portable Web Documents are a clear use case*
- But: current Web Packaging proposal is *not* OPF based
- This may lead to a different packaging in future, also for Digital Publishing

# PWP 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. Content  
  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.  
</p>
```

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

# TECHNICAL CHALLENGE: PRESENTATION CONTROL

Credit: Andrei Niemimäki, Flickr

- 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 PORTABLE WEB PUBLICATIONS

---

- “Portable Web Publications” was, originally, a separate “vision” document (under the name “EPUB+WEB”)
- Was adopted, formally, as part of the group’s work in September 2015, and is now published as an IG document
- The group will contribute to the formulation of the PWP technical challenges, to a better understanding of the requirements
- *PWP is the guiding principle for the group’s further work*

# IDPF, W3C, AND OTHERS

---

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

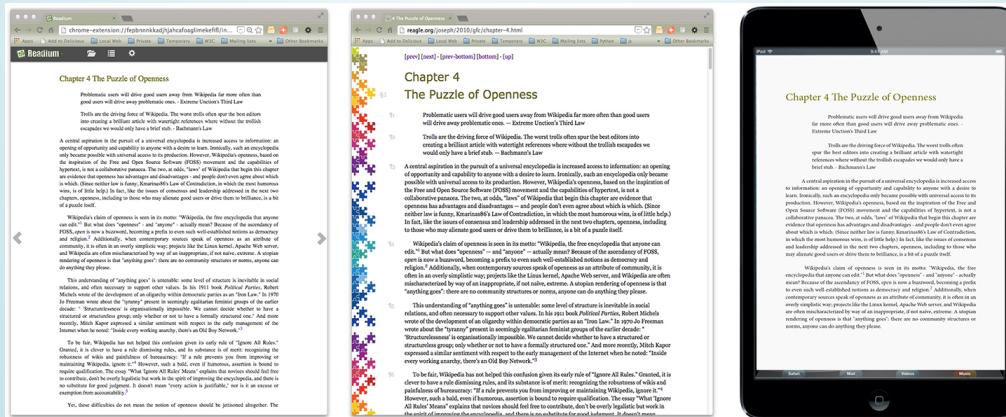
# HOWEVER...

---

- PWP does not replace EPUB 3 (and upcoming EPUB 3.1) at this moment
- Many of the new features will also be part of EPUB 3.1 (e.g., structural semantics)
- The vision is a convergence of the EPUB 3.\* specifications and PWP, eventually

# CONCLUSION

- There is a great potential in a convergence between the Open Web Platform and Portable Web Publications
- 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](https://www.w3.org/dpub/IG/wiki/Main_Page)

## PWP Draft:

<http://w3c.github.io/dpub-pwp/>

## PWP Issue list:

<http://w3c.github.io/dpub-pwp/issues>

## This presentation:

<http://w3c.github.io/dpub/dia-del-w3c-2015-10/index.html> ([PDF](#) is also available for download)

## Direct contact

[ivan@w3.org](mailto:ivan@w3.org)

# SOME REFERENCES

---

## DPUB IG Wiki

[https://www.w3.org/dpub/IG/wiki/Main\\_Page](https://www.w3.org/dpub/IG/wiki/Main_Page)

## PWP Draft:

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

## PWP Issue list:

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

## This presentation:

<http://w3c.github.io/dpub/dia-del-w3c-2015-10/index.html> ([PDF](#) is also available for download)

## Direct contact

[ivan@w3.org](mailto:ivan@w3.org)

**THANK YOU FOR  
YOUR ATTENTION!**