

# 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, “EPUB+WEB”

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

# IDPF AND W3C

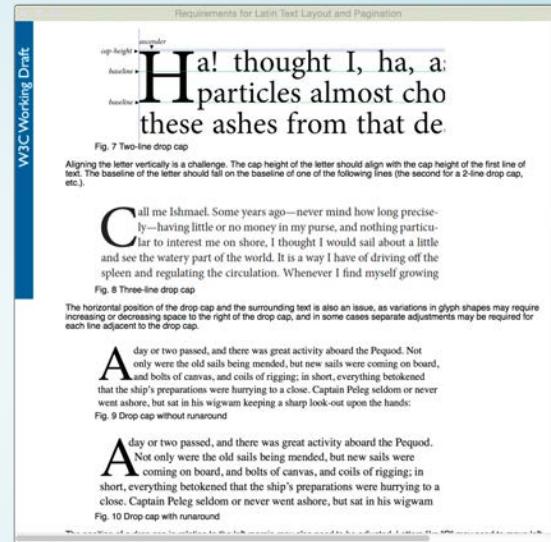
IDPF	W3C
Standards for the Electronic Publishing and Content Consumption (EPUB)	Standards for the General Web Technologies
Builds on lower level Web (e.g., W3C) Standards	Builds on lower level Internet (e.g., IETF, ECMA) Standards
Does <u>not</u> develop standards beyond publishing	Does <u>not</u> develop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

The key is collaboration.

SOME OF THE  
RESULTS OF THE PAST  
TWO YEARS

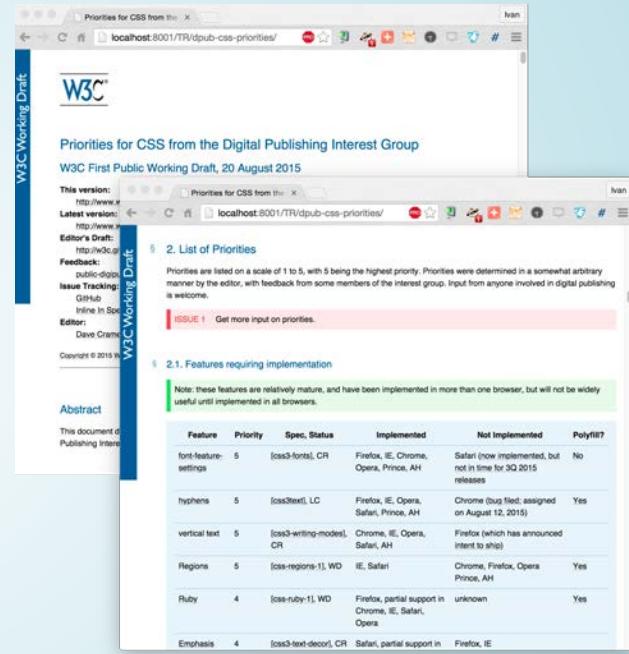
# LAYOUT AND STYLING

- 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 two instances of a web browser displaying the same document. The document is titled "Priorities for CSS from the Digital Publishing Interest Group" and is identified as a "W3C First Public Working Draft, 20 August 2015". The left window shows the main content area with sections like "2. List of Priorities" and "2.1. Features requiring implementation". The right window shows a detailed table comparing CSS features across different browsers based on priority and implementation status.

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 3Q 2015 releases)	No
hyphens	5	[css3text], LC	Firefox, IE, Opera, Safari, Prince, AH	Chrome (bug filed; assigned on August 12, 2015)	Yes
vertical text	5	[css3-writing-mode], CR	Chrome, IE, Opera, Safari, AH	Firefox (which has announced intent to ship)	
Regions	5	[css-regions-1], WD	IE, Safari	Chrome, Firefox, Opera Prince, AH	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

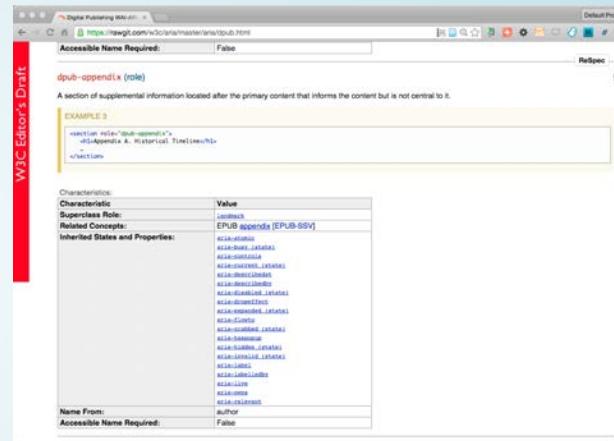
---

- 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

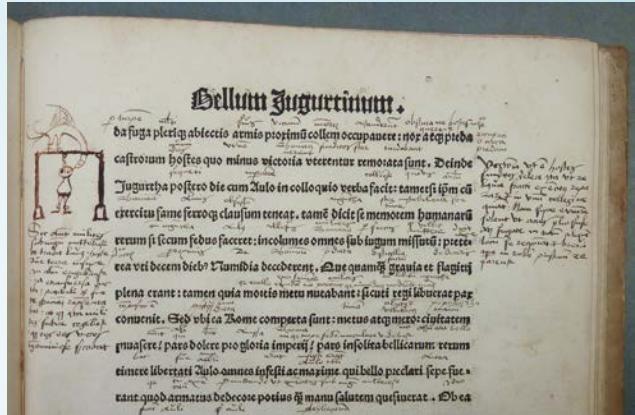
- A Digital Publishing ARIA module is in development
- Publishing terms become part of ARIA

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



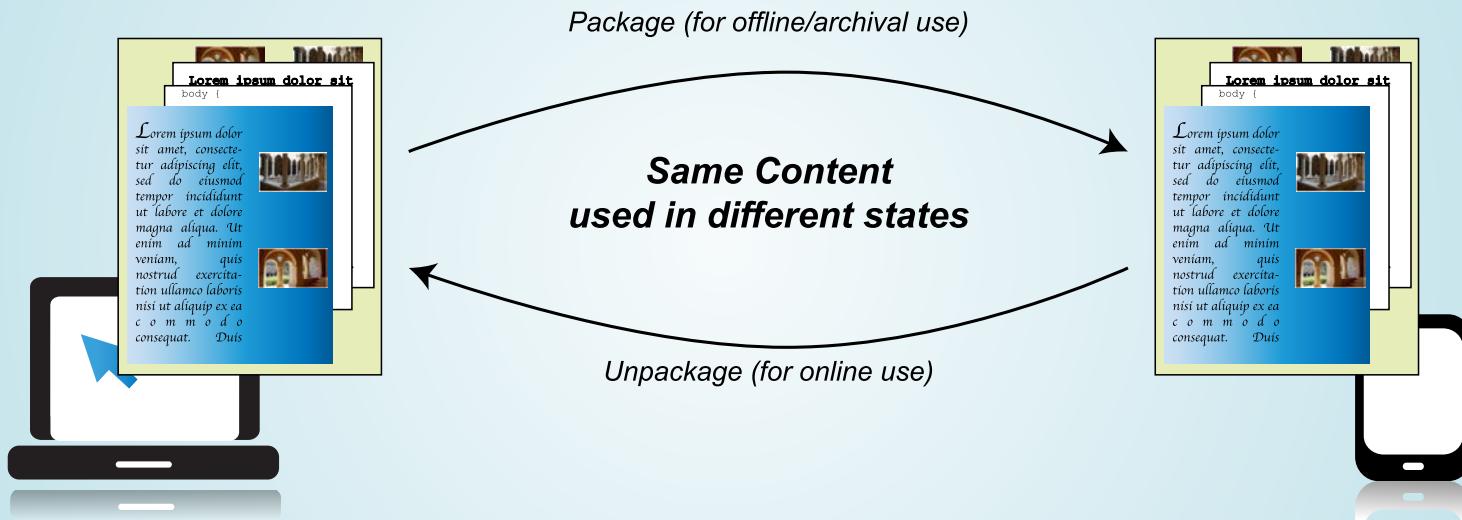
# ANNOTATIONS

- 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



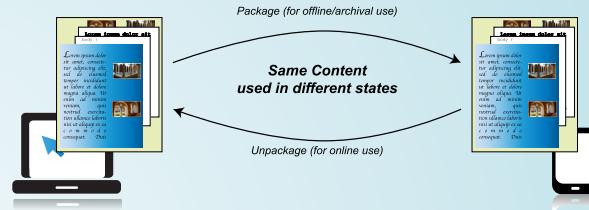
EPUB+WEB

# EPUB+WEB IS A VISION FOR THE FUTURE



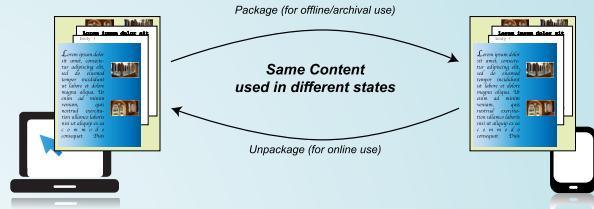
# THE VISION

- Portable Web documents 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

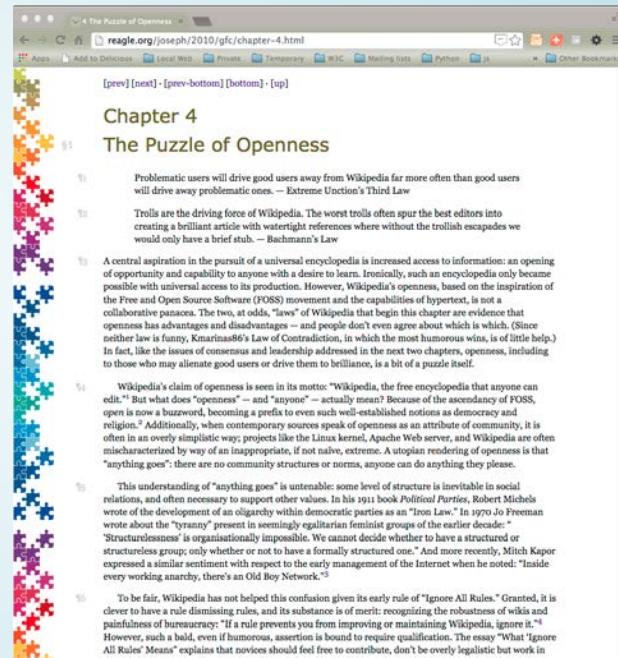
- Publishers can choose to utilize either or both of these publishing modes
- Users can choose either or both of these consumption modes
- Essential features flow seamlessly between on-line and off-line modes, like
  - cross-references, user annotations, access to on-line databases
  - licensing and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



# WHY 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



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...

---

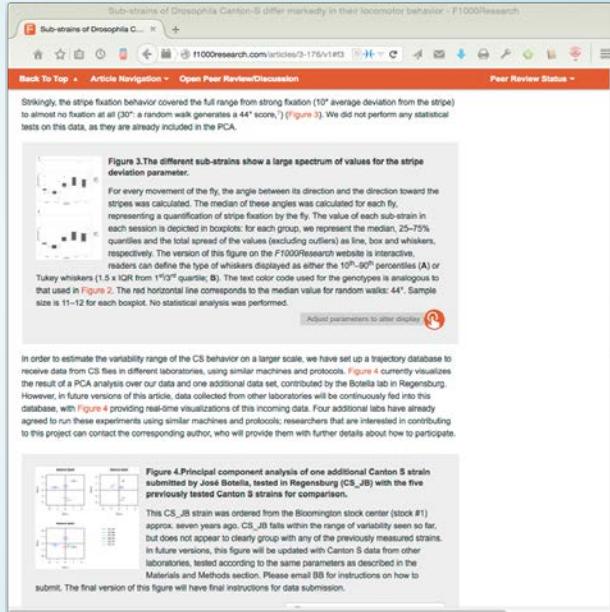
- I may find an article on the Web that I want to review, annotate, etc., while commuting home on a train
- I want the results of the annotations to be back online, when I am back on the Internet
  - Note: some browsers have an “archiving” possibility, but they are not interoperable
    - the content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



Credit: Peliro Agüera, Flickr

# 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!



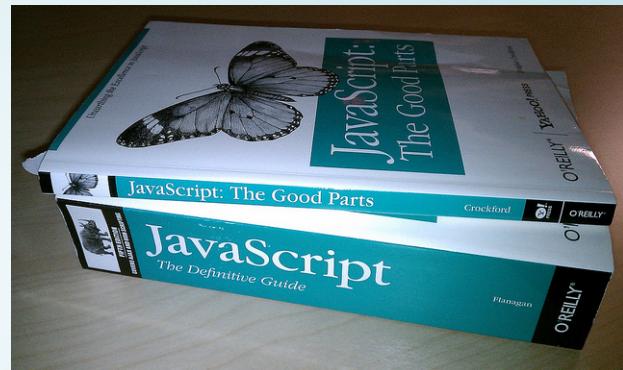
Credit: Merrill College of Journalism, Flickr

# 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

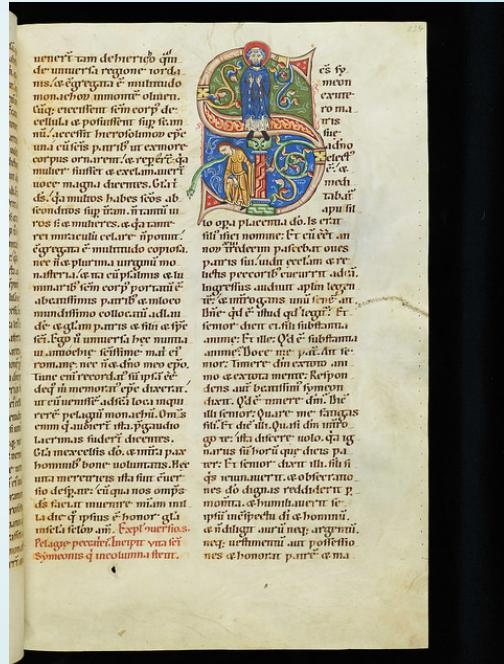
## ADVANTAGE FOR PUBLISHERS' COMMUNITY (CONT.)

---

- 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)

# 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 Digital Publishing in general (e.g., pagination control)
- Some examples follow...

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

# TECHNICAL CHALLENGE: OVERALL ARCHITECTURE

# ADVANCES IN MODERN BROWSERS: 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

# ADVANCES IN MODERN BROWSERS: WEB AND SERVICE WORKERS

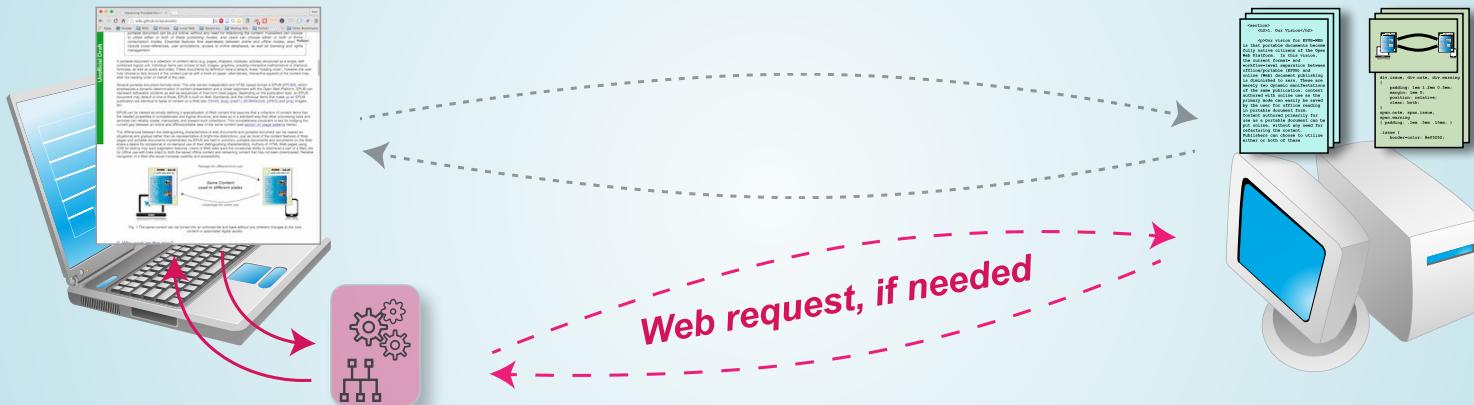
---

- 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

# ENVISIONED ARCHITECTURE: UNPACKED STATE



# ENVISIONED ARCHITECTURE: CACHED STATE



*Requests proxied through Service Worker  
content possibly cached*

# 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

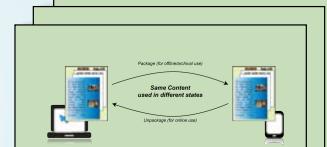
Administrative files  
(manifest, metadata, rights,...)

```
<?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
  - EPUB+WEB is 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

# EPUB+WEB PACKAGING STRUCTURE

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;
```

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

# EPUB+WEB PACKAGING STRUCTURE

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

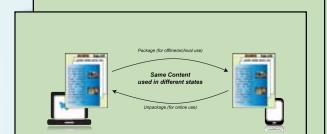
```
{
  "metadata" : {
    "dc:title" : "EPUB-WEB",
    "dc:creator" : [
      "Markus Gylling",
      "Tzviya Siegman",
      "Ivan Herman"
    ],
    "dc:language" : "en-US"
  },
  "manifest" : ...
}
```

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

# 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

---

- EPUB+WEB was, originally, a separate “vision” document
- Was adopted, formally, as part of the group’s work in September 2015
- The group will contribute to the formulation of the EPUB+WEB technical challenges, to a better understanding of the requirements
- *EPUB+WEB is the guiding principle for the group’s further work*

# IDPF, W3C, AND OTHERS

---

- On long term, some EPUB+WEB 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

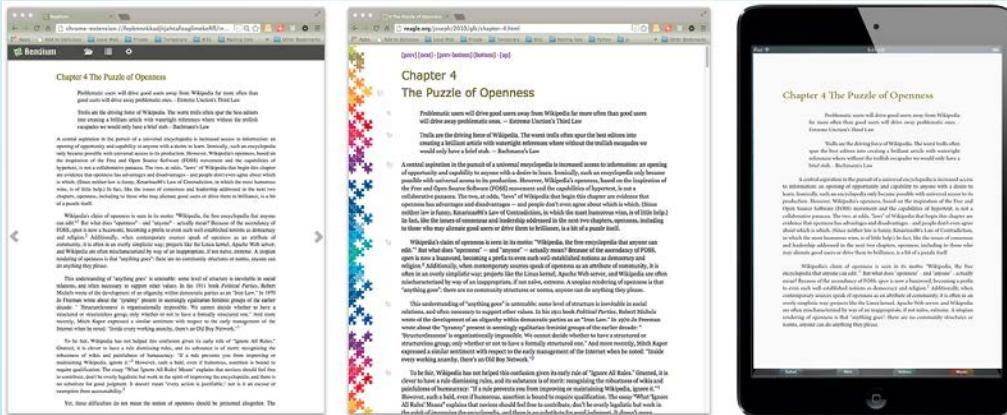
# HOWEVER...

---

- EPUB+WEB 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 EPUB+WEB, eventually

# CONCLUSION

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

## EPUB+WEB Draft:

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

## EPUB+WEB 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!**