

# W3C DIGITAL PUBLISHING INTEREST GROUP UPDATE

Tzviya Siegman, Wiley, and Ivan Herman, W3C

eBookCraft Conference, Toronto, Canada

2015-03-11



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

## 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 annual face to face meetings
  - next in 2015 – semi annual F2F: 26 May in NYC

## DPUB IG 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 our information. We often delegate members to the WGs as well
- DPUB IG has task forces with leads to focus attention on specific issues and collaborate with other WGs

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

**SO, WHO DOES THE  
WORK?**

## EVERYBODY DOES THE WORK!

---

- IDPF has been writing specs for digital publishing for 15+ years, using W3C standards as a basis
- DPUB IG highlights issues in OWP (W3C) that do not support the needs of digital publishing and learns from the W3C
- W3C WGs update specs as needed to reflect needs of digital publishing and the broader web community

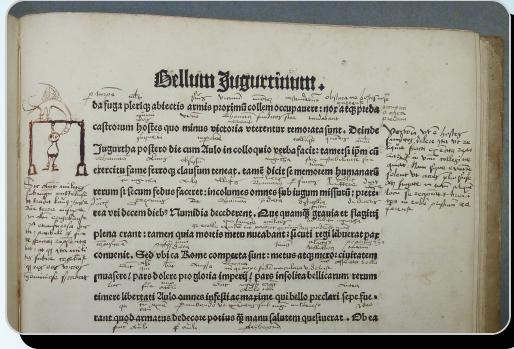
## (CURRENT) DPUB IG TASK FORCES

---

- *Annotations*: led by Rob Sanderson
- *Layout and Styling (aka Pagination)*: led by Dave Cramer
- *Metadata*: led by Bill Kasdorf and Madi Solomon
- *Content and Markup*: led by Tzviya Siegman
- *Accessibility*: led by Deborah Kaplan and Charles LaPierre
- *STEM*: led by Peter Krautzberger

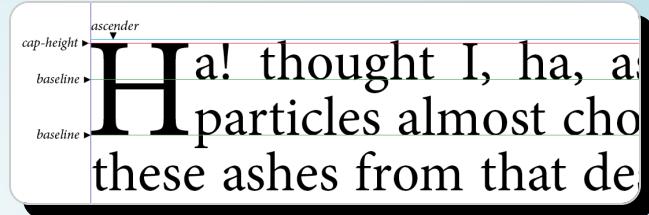
# ANNOTATIONS

- 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



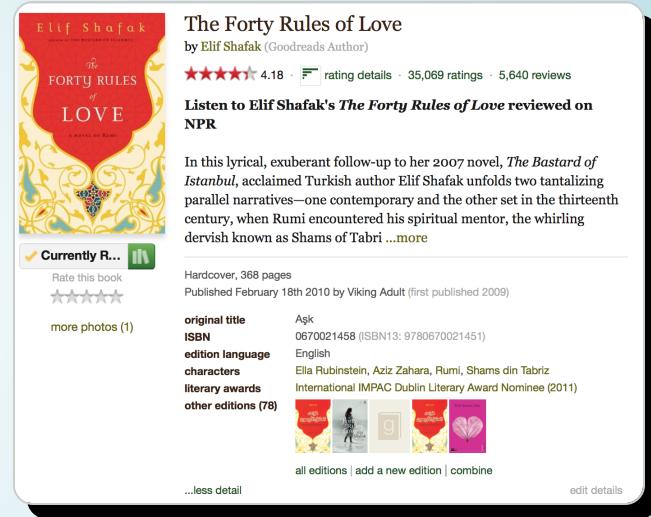
# PAGINATION AND LAYOUT

- Published Requirements for Latin Text Layout and Pagination
  - requirements for print and digital publishing from HTML and CSS using the Latin alphabet
  - not a static document, new releases will come
- Work with CSS WG to publish the [CSS Inline Layout Module Level 3](#), a draft that enabling drop caps and stick-up caps in CSS
- Participation (and gathering [requirements](#)) in the CSS WG's new work standard pagination control
  - as opposed to the current, wildly differing solution used by the industry



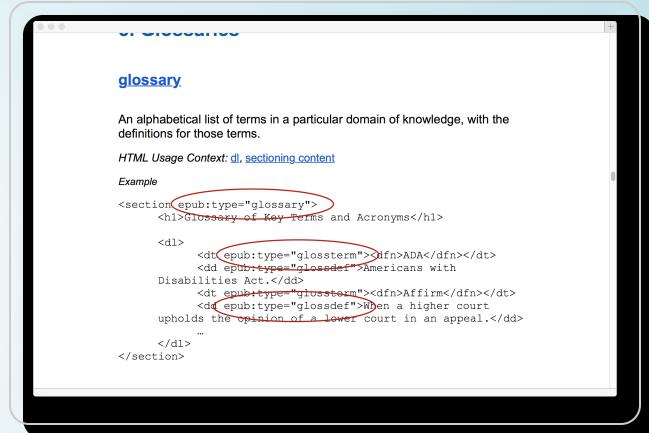
# METADATA

- 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



# CONTENT AND MARKUP

- Goal is to provide a standard way to add a logical structure on top of (but staying compatible with) HTML
  - e.g., identification of various types of sections, index terms, glossaries, etc.
- Working with a W3C Working Group that develops such extensions for accessibility purposes (called ARIA)
- Proposed preliminary list of terms to as a module of ARIA
  - terms provide native HTML elements with specific semantic meaning appropriate to digital publishing
  - working in the context of ARIA provides increased accessibility



# ACCESSIBILITY

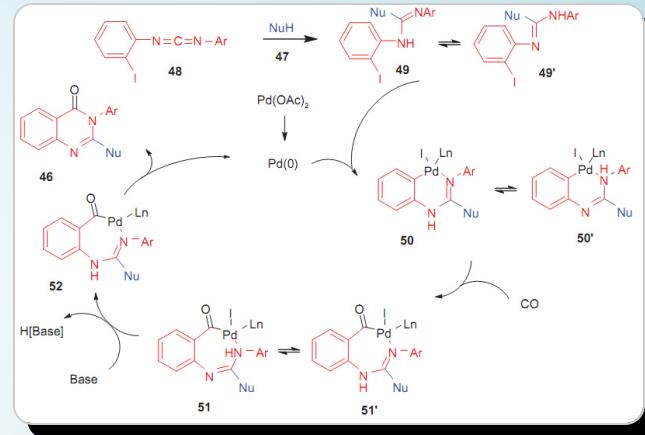
---

- Drafted use cases related to accessibility and personalization in digital publishing
- Reviewing W3C accessibility documentation (UAAG, WCAG, and ATAG) to assess whether the documents reflect the needs of digital publishing and whether publishers are using the documents



# STEM

- Conducted initial interviews with STEM experts to get an idea of major pain points in digital publishing
- Preparing pointed survey to refine information gathered in interviews



## 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

## 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?

The screenshot shows a web browser window with the title "Big Java, Late Objects". The main content area displays a section titled "3.3 Multiple Alternatives". A callout box contains the text: "Multiple if statements can be combined to evaluate complex decisions." To the right, a paragraph explains that in Section 3.1, readers learned how to program a two-way branch with an `if` statement. It notes that many situations involve more than two cases and introduces the concept of multiple alternatives. Below this, it describes an example of a program that displays the effect of an earthquake based on the Richter scale, referring to Table 3.

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).

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.

## THE EPUB-WEB 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

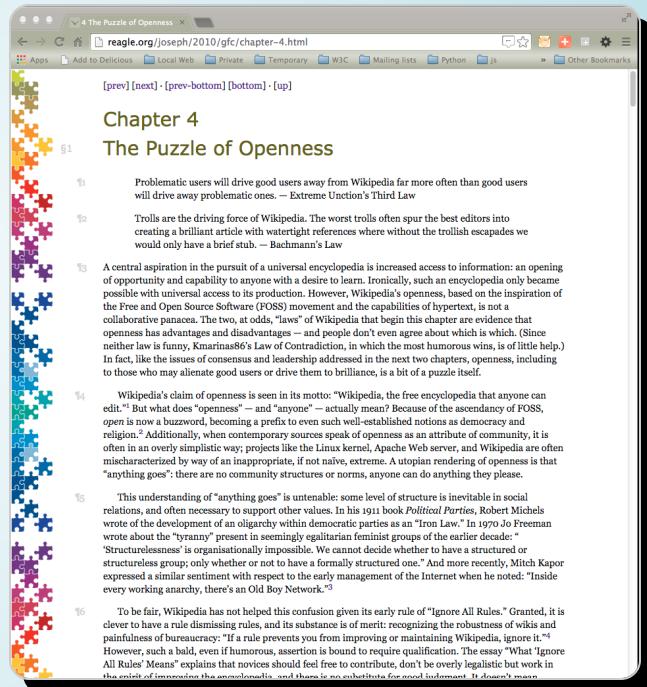
## THE VISION

---

- 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.

# FOR EXAMPLE: BOOK IN A BROWSER

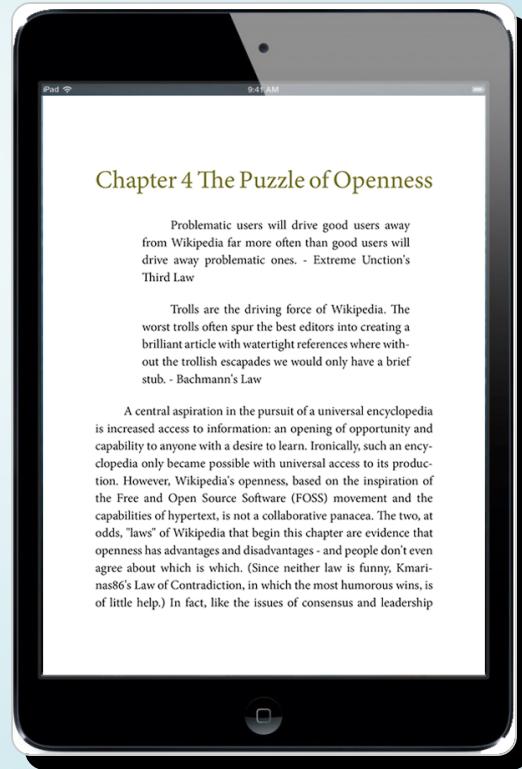
- On a desktop I may want to read a book just like a Web page:
  - easily follow a link “out” of the book
  - create bookmarks “into” a page in a book
  - use useful plugins and tools that my browser may have
  - create annotations



Extract of Joseph Heagles' Book on the Web

## 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)!



## OTHER USE CASES FOR “OFFLINE” AND “ONLINE” USAGE

- Scholarly publishing, including data sets, audio, video, programs, etc.
- In-house publishing: companies (IBM, Boeing, Renault, etc.) publish digitally, with fast refresh time, and need both offline and online
- Archiving and preservation of complete Web documents
- Education materials including “book” content, interactive assessments, animations, a Web client for assessing test results, data for demos, and more



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

## SOME TECHNICAL CHALLENGES

---

- Packaging format better adapted to Web needs (e.g., streaming), see [Packaging on the Web](#) draft
- Document Structure: the current EPUB3 structure (“spine”, etc.) may have to be adapted to Web browsers, with possible defaults and possibility for user interactions
- Identification:
  - general structures for unique ID-s so that documents would really be Web citizens
  - IDs for fragments so that it is possible to link into documents

## SOME TECHNICAL CHALLENGES (CONT.)

---

- Metadata: use of webby methods to identify products, provide means for rich metadata carried with the EPUB-WEB document
- Further improvements on styling and pagination
- Security, access control, and privacy
- Presentation control: reconcile the different traditions of user control (e.g., font size) for documents

## 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

## 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

## UPCOMING EVENTS

---

- 26 May, 2015: DPUB Face to Face @ Hachette Book Group, New York, NY, USA
- 27 - 28 May, 2015: IDPF Digital Book 2015, New York, NY, USA

## 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 White paper:**

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

### **EPUB-WEB Issue list:**

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

### **This presentation:**

<http://w3c.github.io/dpub/ebookcraft-2015-03/> (PDF is  
also available for download)

## DIRECT CONTACTS

---

Tzviya Siegman, Wiley

[tsiegman@wiley.com](mailto:tsiegman@wiley.com)

Ivan Herman, W3C

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

# THANK YOU!