# Publishing and the Open Web Platform

## W3C and the Publishing Industry
## Edupub Conference

Jeff Jaffe, CEO, W3C

# 20 years ago the Web created new experiences for publishing

- ‣ Reading
  - ▪ Hyperlinks  (i.e., non-linear reading)
- ‣ Publishing
  - ▪ Global distribution
  - ▪ Anyone could publish (low barriers)
  - ▪ New advertising opportunities (search engines, pop-ups)
- ‣ But…
  - ▪ impoverished style, layout of early Web no match for print
  - ▪ low resolution screens, slow processors

# Trends of past decade have further transformed reading, publishing

W3C®

- ▸ Internet everywhere
- ▸ Mobility
- ▸ Social
- ▸ Customization
- ▸ Cloud
- ▸ Broadband
- ▸ Multi-function devices
- ▸ Much higher quality display, typesetting, speed

# Many industries feeling the impact

▸ Mobile

▸ Television

▸ Automotive

▸ Health Care

▸ Gaming

▸ Digital signage

▸ Government

# But publishing in particular

W3C®

20% of all US Newspapers are digital

"After 79 years in print, Newsweek goes digital only"

Google Ad Revenue Now More Than U.S. Print Publications Combined

E-Books See Triple Digit Growth As Paper Book Sales Dive

Pew: Survey Finds Rising Reliance on Libraries as a Gateway to the Web

# But publishing in particular

▸ Pew: "**News is becoming a shared social experience** as people exchange links and recommendations as a form of cultural currency in their social networks."

▸ Pew: "In the past year, the **number of those who read e-books increased from 16%** of all Americans ages 16 and older to 23%. At the same time, the number of **those who read printed books in the previous 12 months fell** from 72% of the population ages 16 and older to 67%."

▸ The Bookseller: "In all of 2012, e-book **sales doubled their volume** […] in the United Kingdom"

▸ Pew: "[The] number of owners of either a **tablet computer or e-book** reading device […] **grew from 18% in late 2011 to 33%** in late 2012."
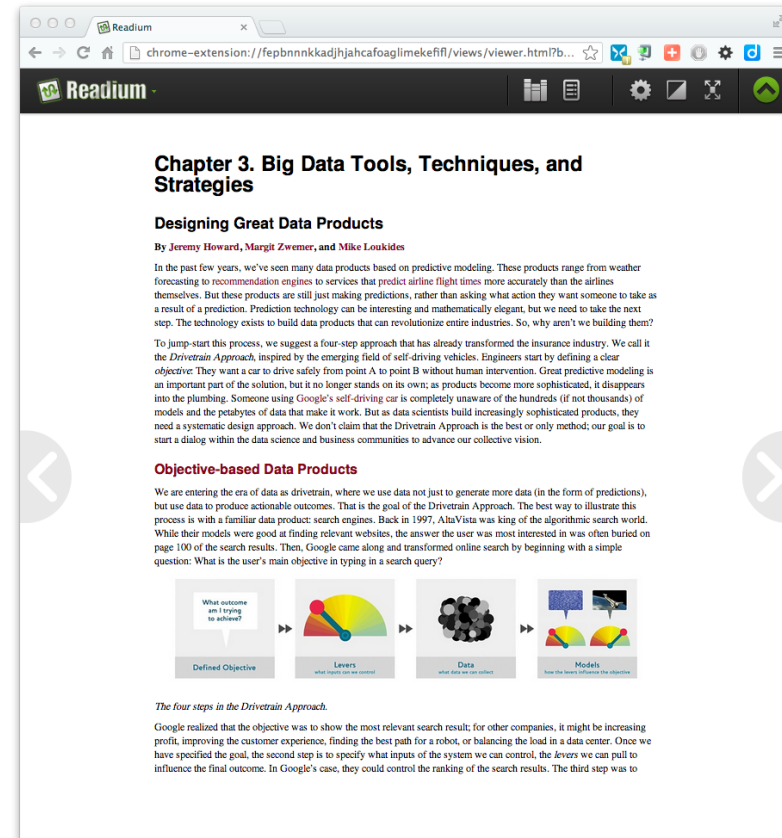
# That is because Publishing = Web

▸ Web is "intimately" tied to the intrinsic purpose of publishing
  ■ the Web's impact on automotive, government, health care, etc., is secondary to their intrinsic purposes.

# What "Ebooks picking up the technology *today*" means

- An EPUB file is, technically, a packaged Web site
  - XHTML 1.1 for EPUB 2
  - XHTML 5 for EPUB 3
- *An EPUB reader is a specialized browser!*

# The Web has become an Open Web Platform

- Web pages are more attractive, interactive and intelligent

- HTML5, SVG, or CSS provides cross-browser and cross-device interoperability

- Video, audio, etc., are a first-class citizens

- Data integration has been simplified

- It has tools for social networking (privacy, security, identity)

- Is the most interoperable platform in the industry

So Publishing will now be further transformed by the Open Web Platform

# Multimedia in News

# Example for an interactive book

# Third-party service integration
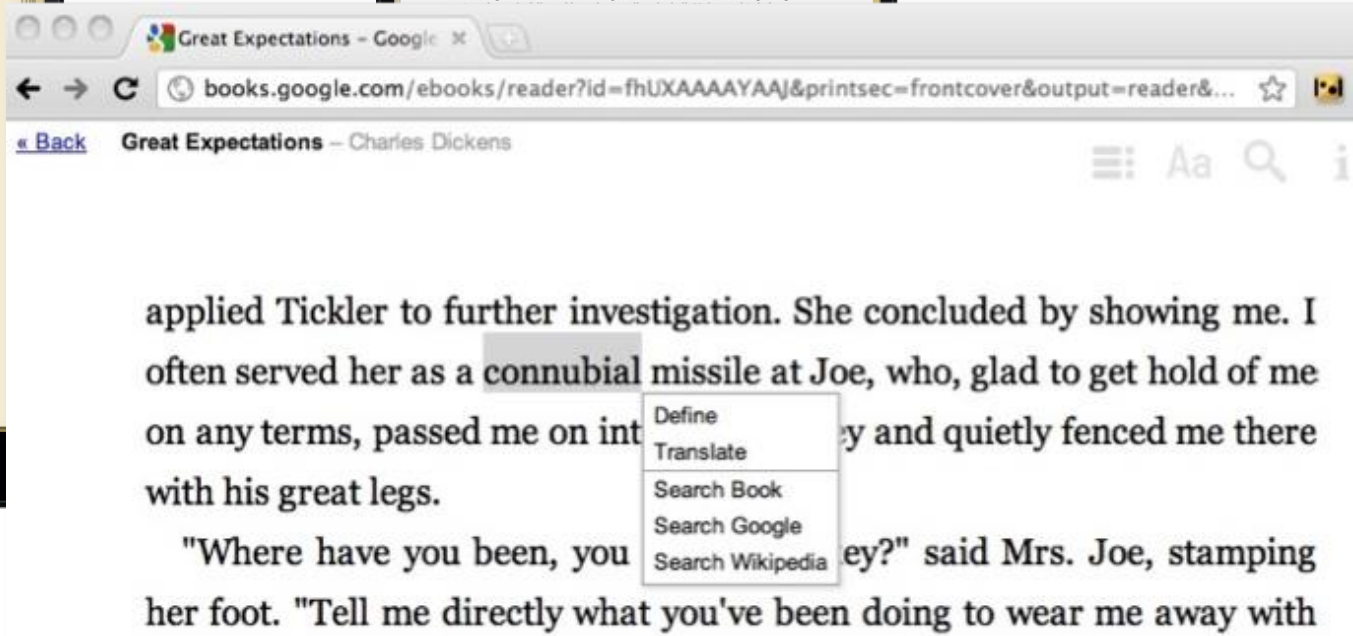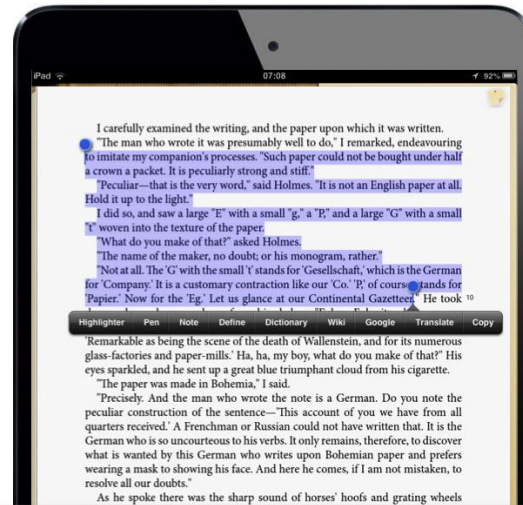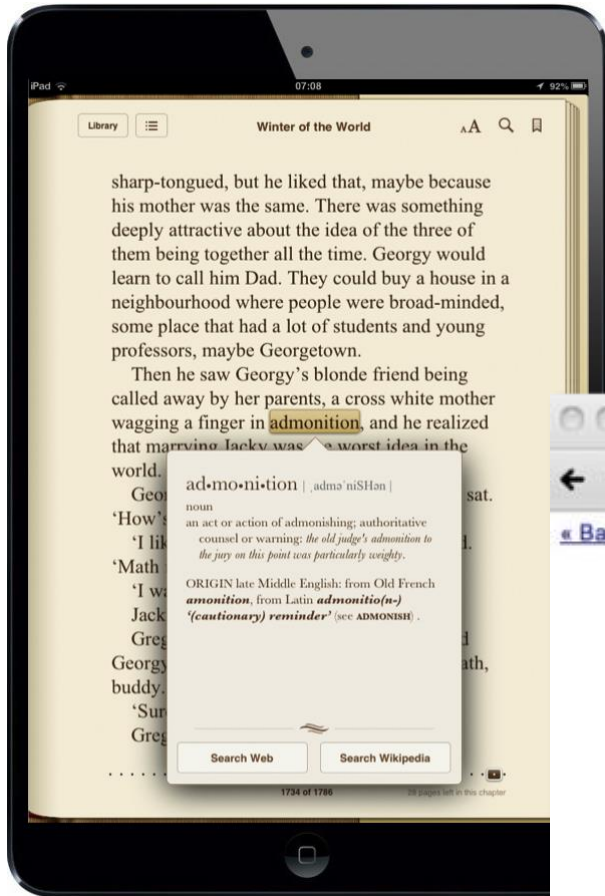
# Social becoming part of reading
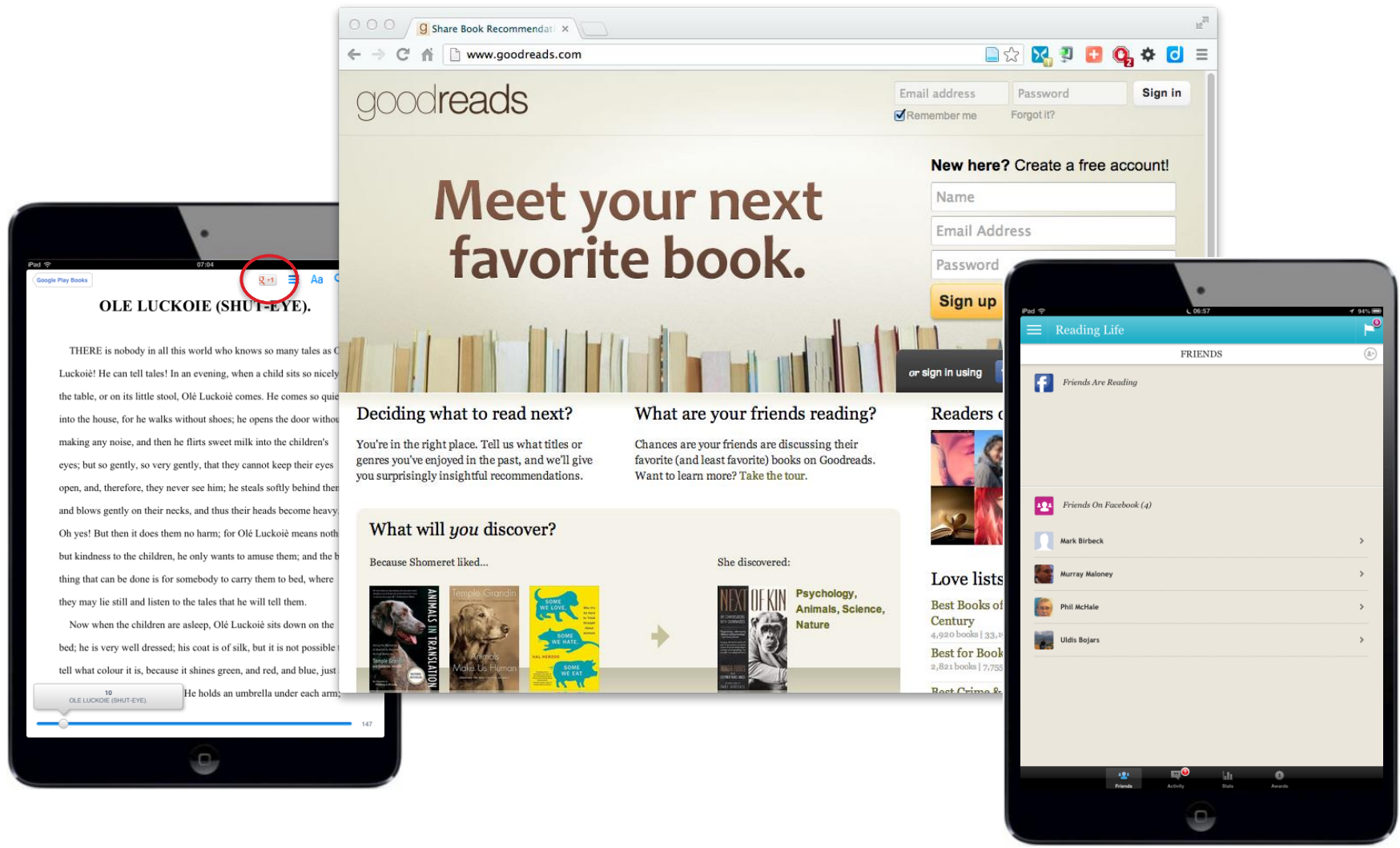
# Inspire new authoring forms (e.g., cell novels)

# Annotating textbooks has always been  part of studying…
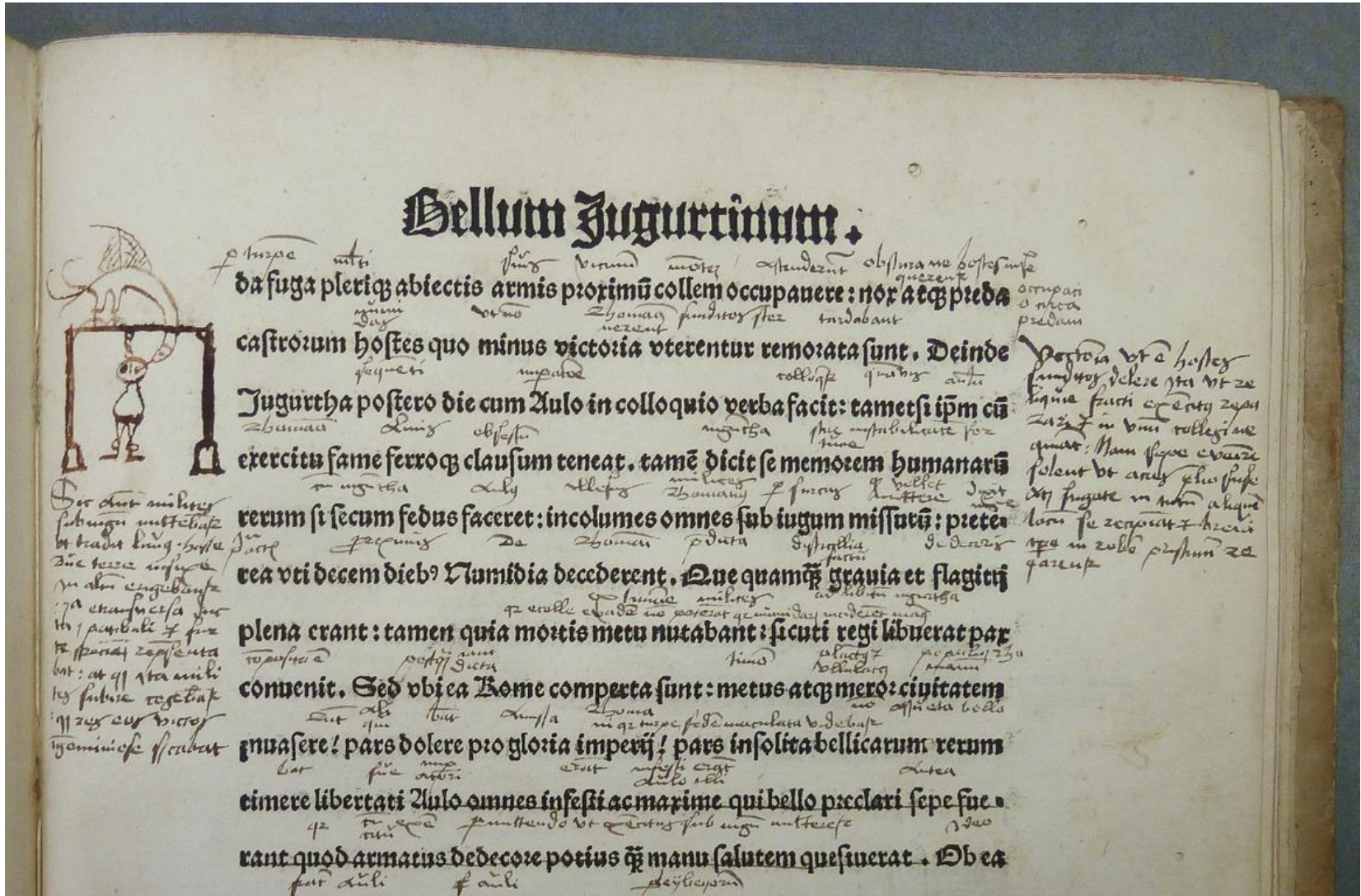
# W3C Open Annotation Community Group

- ▸ Established to reconcile several open annotation models
- ▸ ≈60 members

*Mission:*

Interoperability between Annotation systems and platformsFollowing the Architecture of the WebReusing existing web standardsProviding a single, coherent model to implement without requiring adoption of specific platforms

# And run on any device
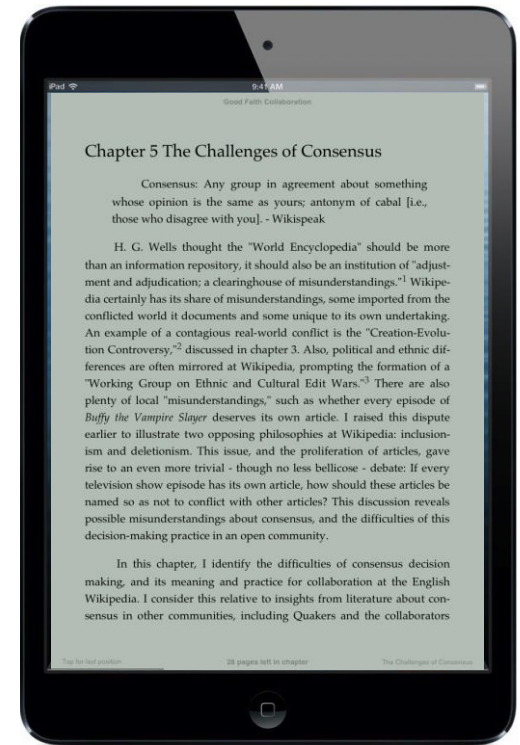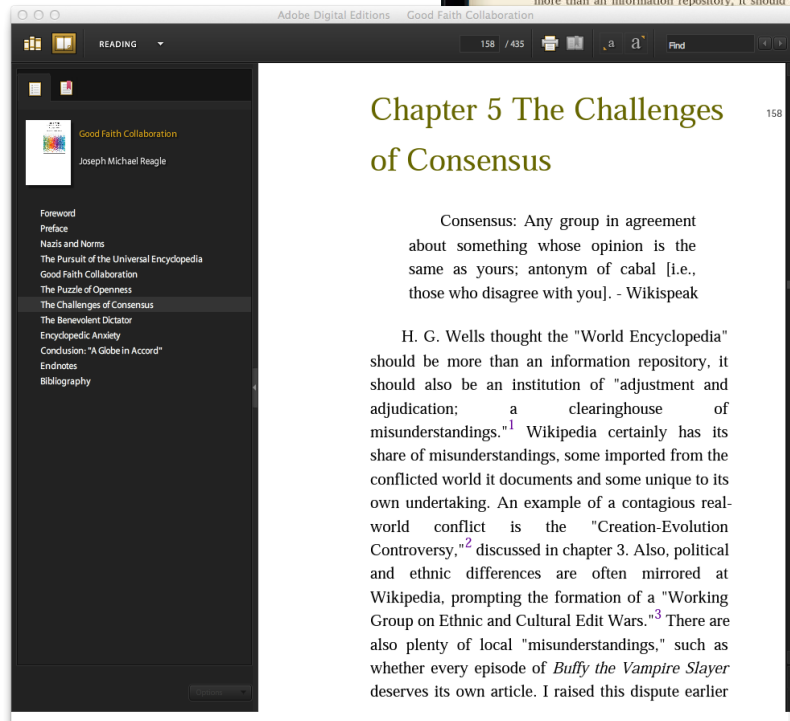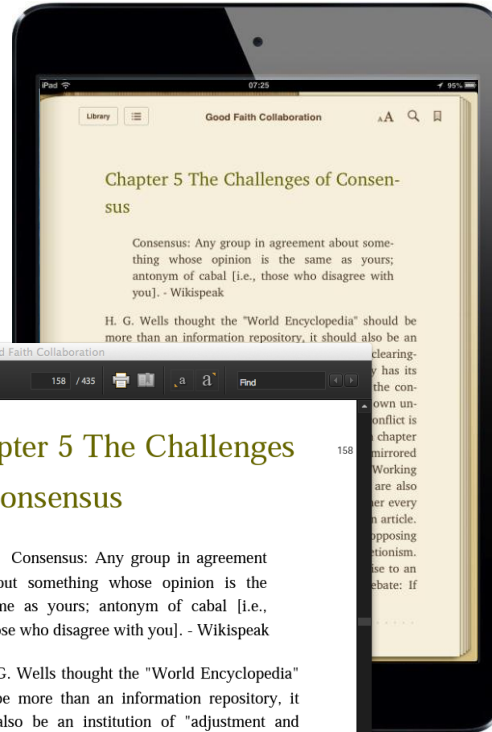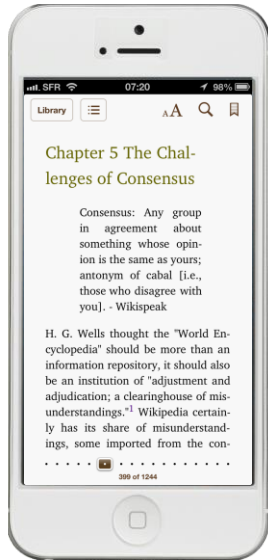
# Is everything perfect and fine?

## a.k.a. are publishers and the Web coexisting perfectly?

# No… ☹

- ▸ Most of the publishing industry players have just been "users" of Web technologies
- ▸ Technology evolution has been driven by "traditional" Web browsing

*The specific needs and priorities of the publishing industry have not adequately been reflected in the evolution of the Web!*

# Some examples of the publishers' needs

W3C®

- ▸ Should match current publishing practices
- ▸ Should leverage the Web at large
- ▸ Should provide more than just text in a file
- ▸ Should support diverse business and distribution models
- ▸ Should satisfy diverse usage patterns

Screen shot: premiumfreebies.eu

*Achieving these will raise technical challenges!*

# "Should match current publishing practices"

- Screens, typography, high quality fonts, colors
- Advanced layout
  - multi-column
  - pagination, headers, footers
  - formatting in the world's scripts
  - hyphenation
  - …
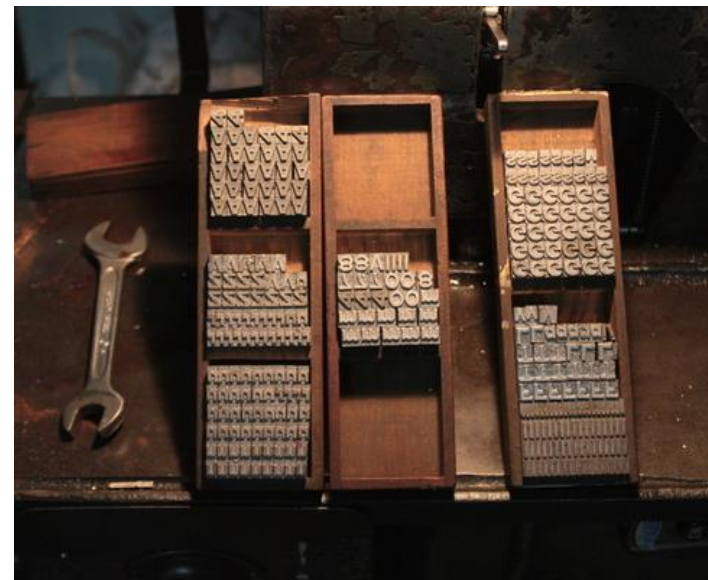
Photo: behance.net

# Example: vertical writing
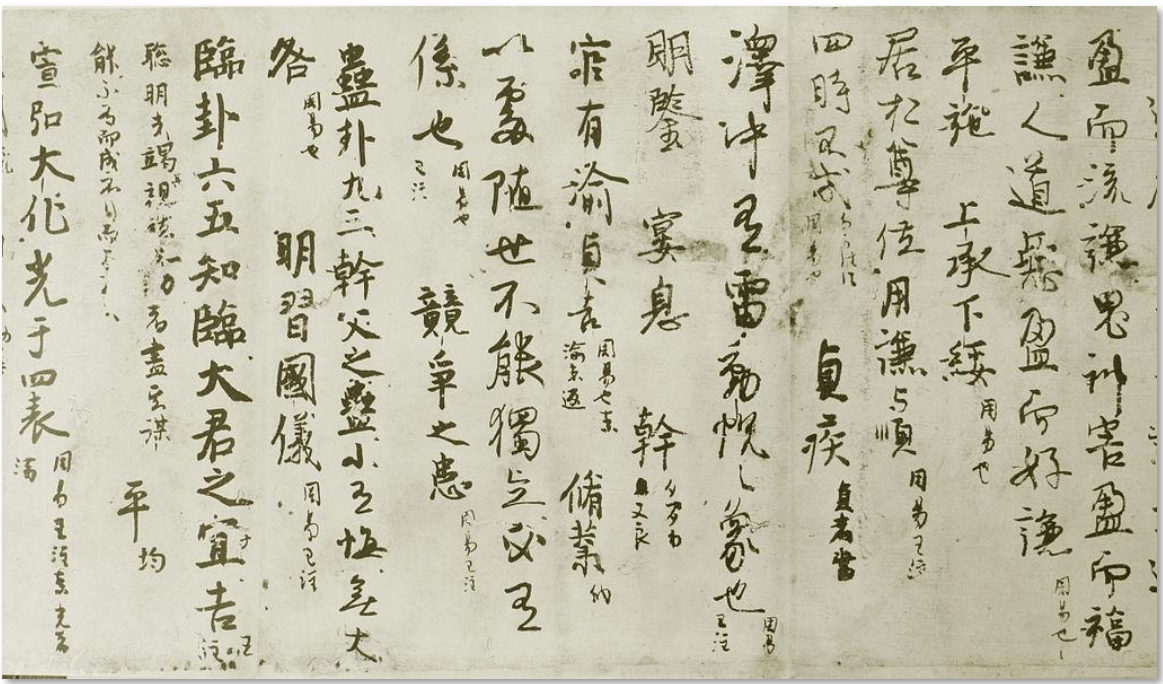


Photo in public domain,
Wikimedia



Photos from "Nemo's great unkle", Flickr

# "Should leverage the Web at large"

▸ Foster sharing, bookmarks, annotations

▸ Data integration, book ids, catalogs, specialized search, discovery

▸ Connect to external services

▸ Facilitate document development (e.g., on-line reviewing, cooperation with editors, on-demand printing…)



From blog.infotrends.com

# "Should provide more than just text in W3C® a file"

▸ Make use of new possibilities

- interaction

- various media

- cross-linking of internal content (indices, lists of tables, etc.)

- high quality graphics

- …



Photo: suttonhoo.blogspot.fr

# "Support diverse business and distribution models"



- ▸ Revenue generation
  - ▪ subscription
  - ▪ ad insertion in eBook apps
  - ▪ social sharing
  - ▪ product placement
- ▸ Web Payments
- ▸ Content protection
  - ▪ one device?
  - ▪ one user?
  - ▪ none?

# Example: book subscriptions
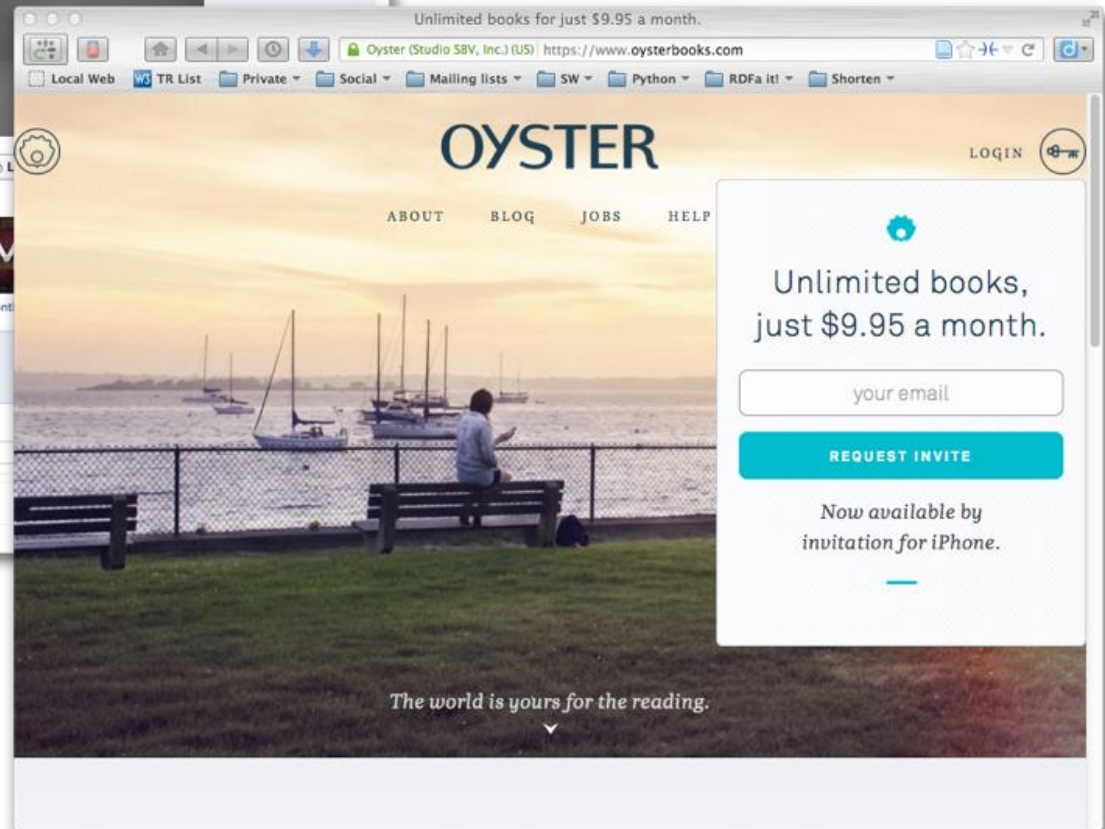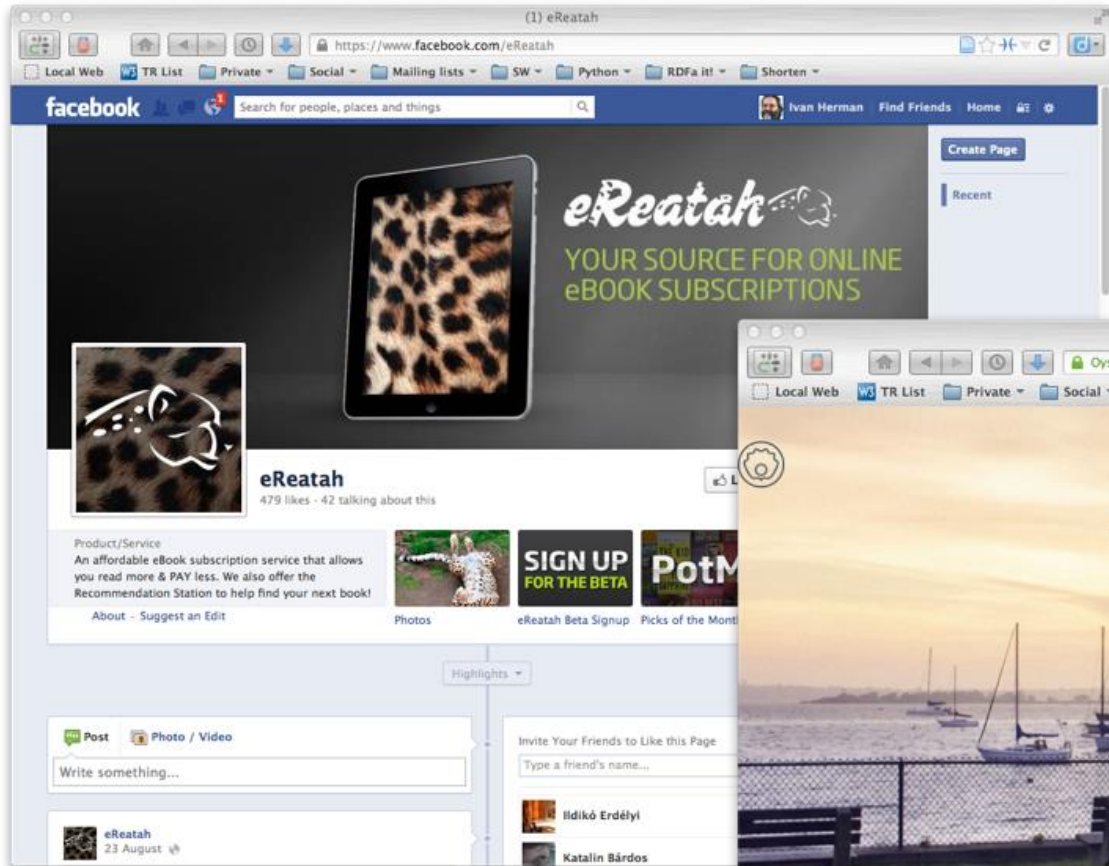
# Example: buying through the reader
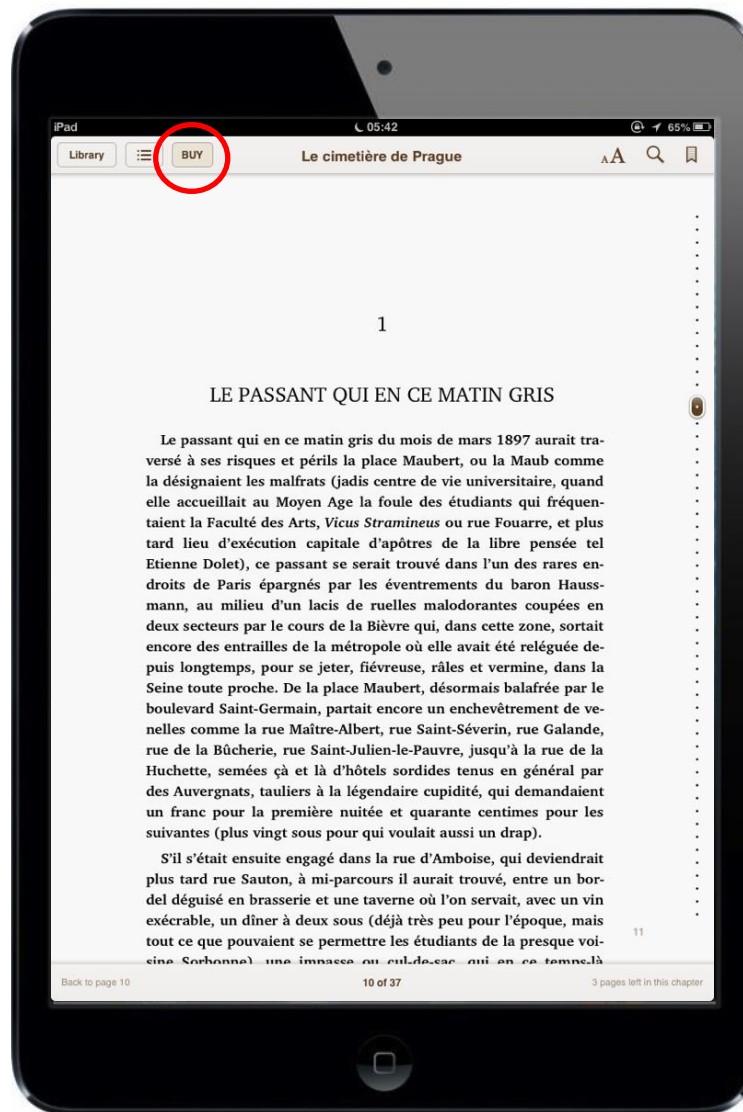
# "Should satisfy diverse usage patterns"

▸ Rich content anywhere, anytime, across multiple devices
  - content portability across devices, readers
  - internationalization, accessibility
  - interaction, graphics, media

▸ Personalization

▸ Inclusion of social features



Photo: extremetech.com

# Example: interoperability

▸ The number of e-reading systems is much larger than browsers'…



From the-ebook-reader.com

# Some interesting OWP challenges in the examples

‣ Digital educational books need
- fixed layout, pagination, advanced typography, hyphenation control
- advanced and interactive graphics (SVG or Canvas object)
- annotation and bookmarking (in a portable manner), local storage
- MathML
- font management
- audio, video
- easy Web payment
- …

# To repeat…

*The technology evolution of the Web has been driven by "traditional" Web browsing*

*The specific needs and priorities of the publishing industry have not been adequately reflected in the evolution of the Web!*

# The way forward: how can W3C help?

# Cooperation

- ▸ Develop cooperation with relevant organizations: IDPF, BISG, IPTC, JEPA, EDItEUR, …
- ▸ Try to be on one another's events, build contacts
- ▸ Create bridges between the developers' communities and publishers
- ▸ *Ensure that the publishers' requirements are channeled to the relevant W3C groups*

# Example: presence in the respective communities' events

▸ W3C keynotes at TOC, IDPF

▸ IDPF presentation at W3C's AC meeting

▸ Joint presence at the CONTEC event in Frankfurt

# Example: increasing cooperation with W3C® IDPF

The EPUB format developed by IDPF is at the bleeding edge of current web standards (HTML5, CSS2.1 /CSS3 modules, SVG…)

## …BUT…

The relevant W3C groups have only weak contacts with the IDPF community,…

**The objective of the IDPF-W3C relationship is to *ensure that the EPUB requirements become central to the OWP developments*.**

# W3C's Workshop series

- "eBooks: Great Expectations", NYC, February 2013 (co-organized with BISG and IDPF)
- "eBooks & I18N", Tokyo, June 2013
- "Publishing and the OWP", Paris, September 2013
- Regional events in Germany, India, …

# Example for results: public list of discussions, presentations

W3C®

other topics. The full list is given in the summary of the wrap up session.

## Main workshop discussions

- Workshop agenda, links to slides and minutes

**First day, February 11th afternoon**
*The first day included: an Overview speech; a first Keynote 1; and two sessions -- the first on Presentation and the second on OWP/ePub. [Minutes from Day 1]*

Following a welcome by session moderator Karen Myers (W3C) and scene setting talk by **Thierry Michel (W3C)** [Slides] the Workshop Co-Chair, **Jeff Jaffe (W3C)** [Slides] presented an overview of the World Wide Web Consortium and the Open Web Platform as a platform for innovation, consolidation, and cost efficiencies. Jeff also presented the expectations for this workshop and the workshop success criteria.

**Bill Mc Coy (IDPF)** [Slides] gave a keynote on the need to increase collaboration between W3C and the IDPF, urging that W3C collaborate on a shared vision and roadmap, building on EPUB 3 as the standard packaged format, for the eBook (portable document) instantiation of the Open Web Platform. Bill also talked about the need to extend current W3C work to address the requirements of the publishing industry, such as high-design content and rich media, accessibility for people with disabilities, internationalization, and semantic structure. He also emphasized the difficulty met by developers when adopting the large number of individual W3C Recommendations that constitute the full Open Web Platform.

**Session 1**
The first Session focused on "Presentation" (CSS, Fonts, etc.). It included four talks:

**Håkon Lee (Opera Software)** gave a demo style presentation showing an implementation of extended layout capabilities in a news magazine format. It demonstrated pagination and multi-column formatting, gesture-based navigation between pages for book-like presentations, all done adding a few lines of CSS to their code. Håkon said that the technical solution he proposed should cover most digital publishing needs.

The second talk by **Vladimir Levantovsky (Monotype)** [Slides] emphasized that digital publications should achieve the same level of typographic quality as print publications. Proposal for topics to be worked on included: high-quality typesetting and font definition and management. To achieve these goals, the standards developed by W3C and IDPF should provide adequate, unified support for all critical technology solutions enabling high-quality typography for eBooks and on the Web.

The third talk by **Jaejeung Kim (Kaist)** [Slides] took the position that electronics books should resemble the paper book in both design and functionality. Do do so he talked about enrichment of eBook user interfaces with a Skeuomorphic approach. He demonstrated a prototype with features such as thumbing-through the book to get an overview concept, or temporal bookmarking by holding a page with one's finger and jumping back and forth between pages, letting go of the bookmarked finger when done.

Finally, **Alan Stearns (Adobe)** [Slides] compared and contrasted the Web and eBook ecosystems. He proposed the adoption of a single solution when interests converge, and where EPUB leads, to improve the Web. He advocated for the prioritization of CSS specs like CSS3-Text and CSS3-Speech. He also talked about Paginated Views or Adaptive Layout as new Web features adapted from EPUB draft (CSS Regions, CSS Exclusions and Shapes CSS Page Templates). Alan also addressed the necessity for testing and the need for ePub readers to contribute their own test cases to, e.g., the Caniuse site.

These four presentations were followed by a discussion and feedback on presentation issues with the audience, moderated by Alan Stearns (Adobe) [Minutes]

**Session 2**
The second session focused on the Open Web Platform and the ePub format and included five presentations:

# Example for results: priority list for W3C® CSS internationalization features

table and caveats related to the table ranking.

| Short name | Description | Experts | Ave | Non-experts | Ave | Weighted score |
|---|---|---|---|---|---|---|
| Vertical text | Vertical text support in CSS needs to be finalised. | 4 4 4 4 3 4 4 4 | 3.9 | 4 4 4 4 3 4    3 | 3.7 | 11.5 |
| Ruby | Ruby markup and ruby styling (especially alignment) needs to be finalized in HTML5 and CSS. | 4 4 4 4 4 3 4 4 | 3.9 | 3 4 4 3 3 3    4 | 3.4 | 11.2 |
| Hyphenation/line-breaking rules | Hyphenation and line-breaking rules for other languages than European ones need to be understood. | 3 4 3 3 3 1 4 4 | 3.1 | 4   2 4 3 3    2 | 3.0 | 9.3 |
| More requirements data | We need to replicate work on the Japanese Layout Requirements initiative for other languages and scripts, and also address special format issues (such as Arabic mathematical layout). | 4 3 3 4 2 2 2 3 | 2.9 | 3 3 3 3 3 3 3 3 | 3.0 | 8.8 |
| Extra long ruby (nakiwakare) | A solution is needed for support of 'extra long ruby' (nakiwakare) when it runs across lines and pages. | 2 1 2 3 4 4 4 4 | 3.0 | 2 3 2   3 3    4 | 2.8 | 8.8 |
| Tate chu yoko | A mechanism is need to support automatic creation of tate chu yoko (horizontal numbers and acronyms in vertical text). | 3   3 1 1 4 2 4 | 2.6 | 3 4 3   3 4    3 | 3.3 | 8.5 |
| Online samples | A way is needed to indicate, in a locale-specific way, which parts of a book should be extracted for samples in online book stores, eg. the illustrations and other information at the start of light novels in Japan. | 3 2 2 1 4 3 4 4 | 2.9 | 2 4 3   3 3    1 | 2.7 | 8.4 |
| Positioning items on page | CSS needs to provide more features for positioning floats in pages, especially vertical centring of items on a page. | 3 3   4 3 1 3 3 | 2.9 | 2   1   3 3    3 | 2.4 | 8.1 |
| Autospace | The autospace feature of CSS that puts visual spacing around embedded Latin text or numbers in lines of ideographic/kana text needs to be supported. | 4 2 3 4 1 1 2 4 | 2.6 | 2   2   3 3    2 | 2.4 | 7.7 |
| Customised line break rules | It should be possible for an author to customise rules for line breaking (eg. kinsoku or geumchik rules). | 3   2 2 3 1 2 4 | 2.4 | 2 4 1   3 3    3 | 2.7 | 7.5 |
| Bopomofo ruby | Tone mark placement for bopomofo ruby needs to be clarified and whatever changes needed to support bopomofo ruby in CSS need to be finalised. | 4 2 3 4 2 1 0 3 | 2.4 | 2   3 3 3 2    3 | 2.7 | 7.4 |
| Switching between vertical & hor | It needs to be possible to switch between horizontal and vertical layouts and automatically make all appropriate changes needed to number formats, punctuation styles, etc. | 2 4 3 2 2 3 1 2 | 2.4 | 3 2 1   3 4    3 | 2.7 | 7.4 |
| Language selection | Selection of language should be based on media queries, not just elements. | 3 2 3 2 1   4 | 2.5 | 3   1   3 3 2 | 2.4 | 7.4 |
| Ruby generalization | Ruby markup needs to be extended to allow for more general annotations (such as for glosses, or to support other languages). | 2 3 2 2 2 3 0 3 | 2.1 | 1 4 3   3 3    4 | 3.0 | 7.3 |
| Rich metadata | Metadata values need to be changed to support markup or annotations (such as for indicating text direction, ruby text, custom embedded fonts for non-unicode characters, language, etc). | 3 2 2 2 2 3 0 3 | 2.1 | 2 3 2   3 3    3 | 2.7 | 6.9 |
| Font availability | There are very few fonts available for use with ebooks (unlike printed fonts), and a significant reason is that font owners are not giving permission for download. | 3 2 3 0 2 1 2 4 | 2.1 | 3 2 3 3 4 4 2 0 | 2.6 | 6.9 |

# New Digital Publishing Interest Group

▸ Co-chaired by Markus Gylling (IDPF) and Madi Salomon (Pearson)

> "The **mission** of the Digital Publishing Interest Group, […] **a forum for experts in the digital publishing ecosystem** […] for technical discussions, gathering use cases and **to better align existing formats and technologies (e.g., EPUB) with the broader Open Web Platform.**"

# In practice…

- The interest group collects use cases and requirements
    - these requirements will be channeled to the CSS, HTML, MathML, SVG, etc., Working Groups
    - members of the IG will also join these groups, if necessary, to develop the necessary technologies
- The interest group contributes to the general OWP interoperability testing effort at W3C
    - ensuring that the special publishing issues are also properly tested by implementations

# It is important that major players join W3C!

▸ Companies outside of W3C can follow what is happening in, say, the CSS Working Group

▸ *But nothing can replace the influence of a company being around the table!*

  ▪ recent members who have joined the work: Pearson, Hachette, …

# Conclusions

▸The future: Publishing = Web

▸Work with W3C work to make this a reality!

Photo by Ivan Herman