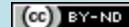


# BRIDGING THE WEB AND DIGITAL PUBLISHING: EPUBWEB

Markus Gylling, IDPF, and Ivan Herman, W3C

BiB Conference, San Francisco, CA, USA

2014-10-23



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

Copyright ©2014 IDPF® and ©2014 W3C® (MIT, ERCIM, Keio, Beihang)

# THE VISION

---

**EPUBWEB is a vision for the future:**

- Portable documents are fully native citizen of the Open Web Platform
- Separation between online (i.e., the “Web”) and portable (i.e., “EPUB”) is diminished to zero

# THE VISION

---

## In other words:

- 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

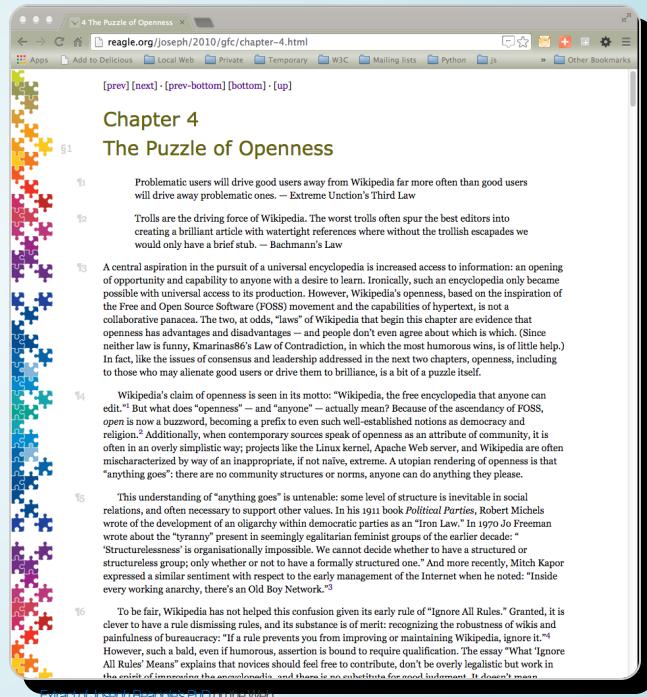
---

- 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?

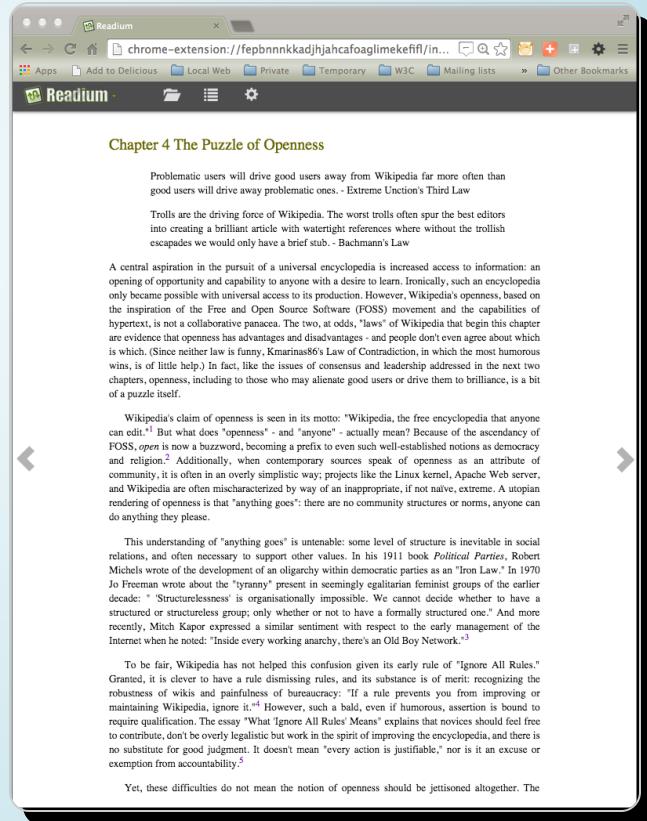
# FOR EXAMPLE: BOOK IN A BROWSER

- 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



# FOR EXAMPLE: BOOK IN A BROWSER (CONT.)

- But:
  - my book may be 2000 pages long
  - conventional Web browsing may not be the right way to view content, a paginated view may be better
  - I may also want to use a small dedicated reader device to read the book on the beach...
- All this should happen using *the same* book, and not a conversion from one format to the other!



This understanding of "anything goes" is untenable: some level of structure is inevitable in social relations, and often necessary to support other values. In his 1919 book *Political Parties*, Robert Michels wrote of the development of an oligarchy within democratic parties as an "Iron Law". In 1970 Jo Freeman wrote about the "tyranny" present in seemingly egalitarian feminist groups of the earlier decade: "Structureslessness is organizationally impossible. We cannot decide whether to have a structured or structureless group; only whether or not to have a formally structured one." And more recently, Mitch Kapor expressed a similar sentiment with respect to the early management of the Internet when he noted: "Inside every working anarchy, there's an Old Boy Network."<sup>43</sup>

To be fair, Wikipedia has not helped this confusion given its early rule of "Ignore All Rules." Granted, it is clever to have a rule dismissing rules, and its substance is of merit: recognizing the robustness of wiks and painlessness of bureaucracy: "If a rule prevents you from improving or maintaining Wikipedia, ignore it."<sup>44</sup> However, such a bald, even humorous, assertion is bound to require qualification. The essay "What 'Ignore All Rules' Means" explains that novices should feel free to contribute, don't be overly legalistic but work in the spirit of improving the encyclopedia, and there is no substitute for good judgment. It doesn't mean "every action is justifiable," nor is it an excuse or exemption from accountability.<sup>45</sup>

Yet, these difficulties do not mean the notion of openness should be jettisoned altogether. The

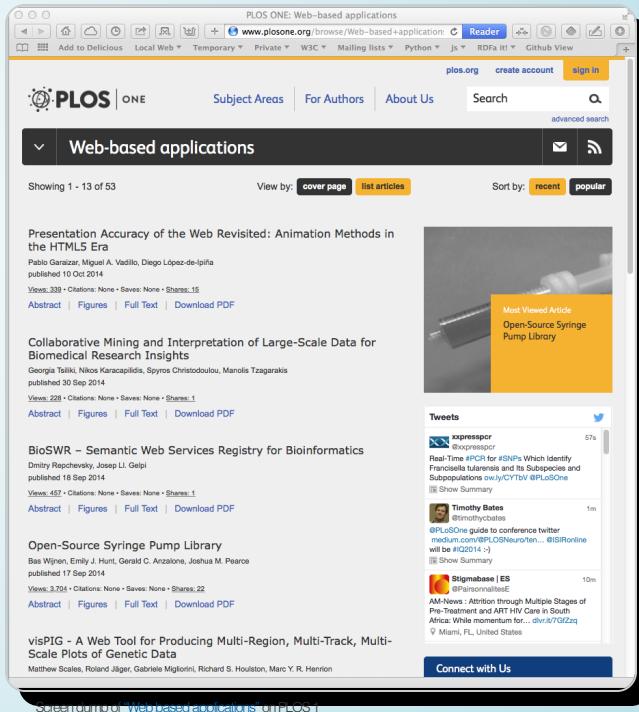
## FOR EXAMPLE: I MAY NOT BE ONLINE....

- 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 no be read on a dedicated reader

Photo credit: Bryan Ong, flickr

# FOR EXAMPLE: SCHOLARLY PUBLISHING

- My paper is published, primarily, online, 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: Pedro Aguiar, Flickr

## FOR EXAMPLE: ARCHIVAL AND PRESERVATION INSTITUTIONS

- Archiving digital assets (i.e., Web pages with *all* dependencies) is a major problem
- There is a need to produce, easily, a *complete* version of a page to be stored through archival facilities

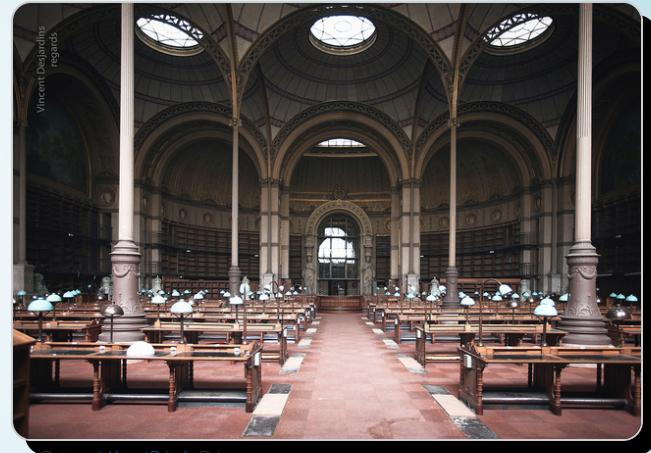


Photo credit: Vincent Dejardin, Milan

## FOR EXAMPLE: EDUCATIONAL MATERIALS

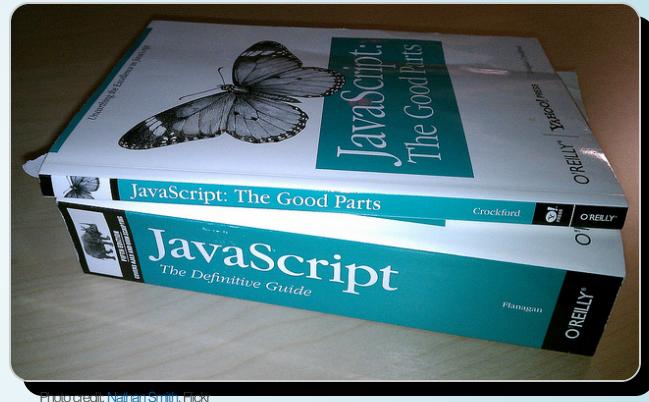
- What is an educational publication?
  - A book of possibly long texts that requires offline access on dedicated devices?
  - A packaged application with built-in interactive tests, animated examples?
  - A Web client reaching out to Web services for assessing test results, to encyclopedia, ...?
  - 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: Menn College of Journalism, Flickr

# SYNERGY EFFECTS OF CONVERGENCE

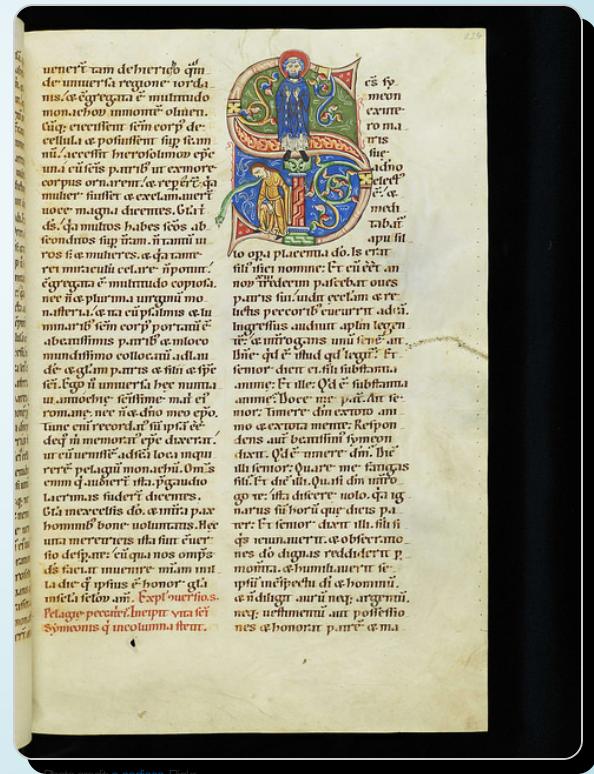
## ADVANTAGE FOR PUBLISHERS' COMMUNITY

- 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: Neilson Smith, flickr

# ADVANTAGE FOR THE WEB COMMUNITY

- Publishers have a long experience in ergonomy, 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 curatory workflow for producing content may become important for Web design

Photo credit: [Eduardo Sánchez](#)

HOW DO WE GET  
THERE?

## HOW DO WE GET THERE?

---

- A strong cooperation between the two communities should be ensured
- Technical challenges must be identified
- A new generation of EPUB (“EPUBWEB”) has to be specified

In what follows *some* of the main technical issues will be highlighted

## ARCHIVAL FORMAT

---

- EPUB is based on ZIP
- There is no standard packaging format for browsers yet...
  - although there is a need for, e.g., applications or data sets
- ... but ZIP may not be the right approach on the Web
  - Multipart Mime may be an alternative
- There is a new work item at W3C on packaging standard, and that may affect EPUBWEB

# OVERALL DOCUMENT STRUCTURE

---

- A complete, offline content may need additional information
  - list of all necessary content, default reading order, etc.
  - in EPUB these are stored in additional, auxilliary files
- But these data may not be necessary for a simple Web page with a few CSS files
  - i.e., some sort of a default structure should be defined
- User interaction paradigms should also be developed to create documents from more complex Web sites easily

# IDENTIFICATION

---

- A concise and unique identification for a *book* is necessary
- I.e., what is the URI for...
  - Shakespeare's Hamlet?
  - its digital edition published by Publisher XYZ?
  - the copy I own and annotate?
- This is necessary to make a book a first-class citizen on the Web
- This is already the topic for huge debates in the publishing and library community...

## IDENTIFICATION (CONT.)

---

- Unique identification is not enough
- A fragment identification framework is also necessary to link *into* the book
- There are fragments defined for various media, but a universal approach, workable for browsers, is still missing
  - these should be agnostic to offline vs. online state, to media type, etc.

## IMPROVEMENT ON STYLING, PAGINATION

---

- Books usually need higher quality typesetting than average Web pages
  - control over fine aspects of drop caps, alignments, justification, hyphenation, etc.
  - cf. <http://www.w3.org/TR/dpub-latinreq/>
- The main issue, however, is *paged view* of the content; what is a “page” for an electronic content?
  - Is this a new CSS concept? Do we need an extension to the DOM?

# 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 may be more user control
    - foreground/background color
    - choice of fonts
- There is a need to reconcile these traditions

# 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!



# WHAT IS NEXT?

## WHAT IS NEXT?

---

- We collect comments on this vision
  - this conference
  - F2F meeting of the Digital Publishing Interest Group next week
  - internal discussions at IDPF and W3C
- Plan to release a public White Paper in November
  - taking into account all the comments we receive
- Extensive discussions with various communities should then happen

## IF THERE IS CONSENSUS

---

- Work with existing IDPF and W3C groups, where necessary, on specific details
- Set up a new group (or groups) to define the EPUBWEB specific issues
  - lots of details to be clarified...

---

# THANK YOU FOR YOUR ATTENTION

These slides can be accessed on the Web (in HTML):

<http://www.w3.org/2014/Talks/1023-SF-IH/>

Markus Gylling, IDPF

[mgylling@idpf.org](mailto:mgylling@idpf.org)

Ivan Herman, W3C

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