The Future of Standards for Digital Publishing

Liza Daly, Markus Gylling, and Ivan Herman
IDPF Digital Book Conference 2014
2014-05-29
Why is W3C interested in this area?
We know that...

• …digital publishing formats rely heavily on W3C’s core technologies, i.e., the Open Web Platform (OWP)
  • (X)HTML5, CSS 3, SVG, MathML, SMIL, scripting…

• …publishing workflow systems are also moving towards OWP (O’Reilly’s Atlas, Chaucer, Inkling, …)

• …publishing represents one of the biggest use cases for the OWP after “traditional” browsers
Most of the publishing industry players have just been “users” of Web technologies
The technology evolution of the Web has been driven by “traditional” Web browsing.

The specific needs and priorities of the publishing industry may not be reflected in the evolution of the Web!
There are of course some successes
We got some decent portability
We can produce books with nice graphics and with fixed layout...

Webb will be the largest telescope ever placed in orbit. To get this huge telescope into space, it will have to be folded up inside the rocket that carries it from Earth. It will unfurl as it nears its destination, the mirror unfolding and the sunshield spreading like wings.

With its infrared vision, Webb will be able to see light from vast distances that has been stretched as it travels across the expanding fabric of space – a process known as “redshifting.” This will enable scientists to see the light from the first galaxies to form in the early universe – the first flicker of stars that formed after the Big Bang.

Hubble has seen the adult, teenage and child galaxies of the universe, but Webb will see the cosmos’ toddlers and infants.

Webb’s infrared view will allow us to see through opaque clouds of gas and dust in our own galaxy to objects within, like newborn stars and dusty disks forming new solar systems, expanding and deepening our knowledge about how stars and planets develop.

And Webb’s instruments will be able to analyze the light from planets around other stars, allowing us to search for traces of water vapor that could indicate the presence of life-giving oceans. Webb’s vision will be able to detect features in the atmosphere of a planet that could show biological activity.

Webb’s launch is scheduled for 2018. And despite the high expectations for the telescope, the true excitement will come from the discoveries Webb makes that no one planned. When Webb takes its place among the stars, it will give astronomers an unprecedented tool to explore the cosmos. The greatest science it reveals may – as with Hubble – open avenues to questions astronomers have not yet thought to ask. The true power of Webb may be its potential for unbounded, unexpected discovery.
...that can also adapt themselves
We got some level of vertical writing sorted out
But everything is not perfect...
Widow control possibilities are incomplete

Sacrifice of Isaac

other man’s charms that he had no time to think about the fact he was lying. Briefly he let the image of Chevejon back into his mind: The strong face under its thick bangs exuding a confidence that was virtually impossible not to trust. It was extraordinary: That you recognized him as a master hustler, that you were scared of him, did nothing to diminish your urge to trust him. To be ascendant over such a man, what could Danni be?

Could he be trusted? Luke yawned, drained now, tiredly recognizing the temptation. Of course he could not be. Still, he was all Luke had.

The twelve hours to Chevejon’s return dragged interminably. He ordered breakfast, ate, tried to marshal his thoughts. But he could not concentrate. Now, for the first time, he realized that Chevejon had not told him a thing, not a single thing, about his brother. He had revealed nothing. As if it were a lesson in memorization, he reminded himself that he must not trust Chevejon. And that he must be aware of what Chevejon might have in mind for Luke. At this thought, fear surfaced; he closed the bedroom curtains, stretched out on the bed, and fell asleep.

Chevejon spent less time with me during that day, leaving me—to my disappointment—in Giorgio’s care. The morning he spent in his office. There was much to think about.

For one thing, he now had two children, not one, to baby-sit. Giorgio was going to have to be put on retainer; he needed both his bodyguards for the boy. Overkill was in order here, for he had to be very, very careful. And as for him, he saw quickly, a different kind of pressure was going to be required. The boy must be afraid, that was sure; he must think that Chevejon was prepared simply to toss him away. Which was easy, because it wasn’t far from the truth. He knew the little esteem in which Danni held his family; he did not expect that his partner would object to anything he did to dispose of the absurd threat this boy posed.

But then, before he did dispose of him, he needed better to understand exactly what that threat was. This was not a question that could be posed directly; as with the girl, there was nothing to be gained by directness. So, in fact, he was cornered.

Sitting in his office, unfazed by his lack of sleep, he thought carefully. No, he was not concerned; there was something he could use with the girl. It was about time this cat-and-mouse game with her had its end. And if he could not pose a question directly, he could a threat.
January 4

Pennsylvania city, each of her days could easily resemble the next—she has a lot of time, having been laid off from her job busing tables at a fast food restaurant. She has enough money to live on, as a recipient of government assistance for people with disabilities.

But Beth also has something else: ingenuity.

This trait isn’t generally ascribed to people living near the outskirts of society’s citizenry.
January 4

Pennsylvania city, each of her days could easily resemble the next—she has a lot of time, having been laid off from her job busing tables at a fast food restaurant. She has enough money to live on, as a recipient of government assistance for people with disabilities.

But Beth also has something else: ingenuity.

This trait isn’t generally ascribed to people heading up the social services of a city’s citizens.
We have to understand the requirements of various reading systems.

\[
f(x) = \begin{cases} 
\sum_{i=1}^{s} x^i & \text{if } x < 0 \\
\int_{1}^{x} x^i \, dx & \text{if } x \in S \\
\tan \pi & \text{otherwise (with } \pi \approx 3.141) 
\end{cases}
\]
Publishing needs advanced features

- Advanced and interactive graphics
- Internationalization (Ruby, Bopomofo, vertical writing,...)
- MathML
- Font management
- Fixed layout

- Audio, video
- Headers, footers, ...
- Bridging on-line/off-line
- Harmonization of metadata
- Accessibility features (e.g., speech/braille publications)
- ...
Publishing needs advanced features

- Advanced and interactive graphics
- Internationalization (Ruby, Bopomofo, vertical writing, ...)
- MathML
- Font management
- Fixed layout

- Audio, video
- Headers, footers, ...
- Bridging on-line/off-line
- Harmonization of metadata
- Accessibility features (e.g., speech/braille publications)
- ...

Most of these are being worked on at W3C, but still need extra care and clear requirements!
What can be the Role of W3C?
It is time for the Publishing Industry and the traditional Web community to cooperate in defining the Open Web Platform!
How to achieve that?
Creation of a “Digital Publishing Interest Group”

- Co-chaired by Markus Gylling (IDPF) and Liza Daly (Safari Books Online)

“The mission of the Digital Publishing Interest Group, [is to be] a forum for experts in the digital publishing ecosystem [to have] technical discussions, gather use cases and to better align existing formats and technologies (e.g., EPUB) with the broader Open Web Platform.”
In practice...

• The Interest Group has defined a series of Task Forces
  • the task forces collect use cases and requirements
  • these are, eventually, passed on to the relevant W3C groups (HTML, CSS, etc.)
• Reviews, from the publishers’ point of view, other groups’ relevant draft specifications
In practice...

- The group consists of representatives of the publishing industry
  - Pearson, Hachette, IDPF, Apple, Adobe, Benetech, Nokia, Monotype, BISG, …
  - further discussions with others prospective members…

- Some IG members are also part of other W3C groups, working HTML, CSS, MathML, etc.
Some examples for concerns addressed by the W3C Group
Books that require rich media...

- Requirements:
  - high quality graphics, video, interactive graphics
  - complex layout (text flowing through different part of the page)
  - user specific notes, bookmarks, annotations
... but with high quality aesthetics and ergonomics

- Requirements:
  - control over headers, footers, running elements
  - control over widows, orphans, hyphenation, justification
  - fixed or reflowable layout
  - high quality graphics, both in terms of bitmap images as vector graphics
  - drop caps
  - font management
• The **DPUB IG Layout & Styling Task Force** addresses these fundamental rendering issues
  • text flows, drop caps
  • hyphenation, widow, orphan, etc., control
  • video, audio
  • page layout with headers, footers, running heads

• These requirements are translated into specific requirements to be taken up by the CSS, HTML5, and SVG Working Groups
STEM publishing

- Requirements:
  - font management
  - high quality graphics
  - mathematical formulae
  - extensions to HTML for specific content (e.g., chemical formulae)
  - separation of “content” from explanations, notes, etc.
  - powerful indexing, table of contents
• The **DPUB IG Content & Markup Task Force** identifies ways to add structural information to the raw content:
  • terms to be indexed; sections that are used for notes, sidebars, explanations; accessibility information

• The **DPUB IG STEM Task Force** has collected a number of STEM specific use cases
Annotations

- Annotations belong to the traditions of reading
- Annotations can be private, can also be shared among friends, colleagues, ...
  - classroom usage in education
  - sharing notes among friends on a new publication
- Annotations can be text, drawings, different types of media
• The **EPUB IG Annotation Task Force** has collected a number of eBook specific use cases and requirements

• The **W3C Open Annotation Community Group** has defined a general data model for annotations

• W3C may start an **Annotation Working Group** to define a standard, interoperable environment to create, store, and share annotations
Search and discovery

- Information on books should be discoverable on the open Web
  - basic information
  - reviews
- Not only on one, proprietary site…
• **DPUB IG Metadata Task Force** looks at issues on making books discoverable on the Web
  • what type of metadata is appropriate for, e.g., search engines
  • how to express those
  • what vocabularies to use
These are just examples!

- Lots of other issues may be discussed
  - the “pain points” are numerous
- We need a larger participation of the publishing industry
  - W3C cannot do it all by itself!
Why this matters to you.
EPUB 3 adoption is slow
EPUB 3 adoption is slow

Chicken & egg problem persists:

Not enough EPUB 3 content to support.

Not enough support to create EPUB 3 content.

Few reading systems on epubtest.org score > 75%
How did we end up here?
Reading systems: We’re lazy
Reading system developers would much rather differentiate on custom features than re-implement EPUB 3 from scratch each time.

Pushing features into the open web lets us get back to innovating.
Publishers: You’re busy
Publishers can’t invest in massive QA efforts just to make use of advanced features.

Advocating for adoption of web-based ebook standards lets you get back to publishing.
Towards a network of books
Towards a network of books

The most desired features require cooperation among reading systems, search engines, & browsers:

- Portable annotations
- Social sharing & commentary
- Discoverability
Towards a network of books

The most desired features require cooperation among reading systems, search engines, & browsers:

Portable annotations
Social sharing & commentary
Discoverability
(Monocultures are dangerous)

Currently Unavailable

Want us to email you when this item becomes available?

Email me
What publishers can add
What publishers can add

What changes to HTML could trivially improve ebook markup?
What publishers can add

What’s the best way to present textual content that promotes readability, comprehension, and retention?
What publishers can add

What’s the best way to present textual content that promotes readability, comprehension, and retention?

What cultural traditions in reading and writing should be brought to the open web?
Help books become natural citizens of the open web
Properties of a naturalized OWP citizen

“The mission of the Digital Publishing Interest Group, [is to be] a forum for experts in the digital publishing ecosystem [to have] technical discussions, gather use cases and to better align existing formats and technologies (e.g., EPUB) with the broader Open Web Platform.”
bridging online and portable publishing

- today, online and ebook publishing are separate forms
  - typically separate channels, separate clients, separate experiences

- portable document as a state, not as a form means that:
  - an online document can change state to portable at user request
  - ... while retaining all functionality
  - ... the publisher can choose which of these states to enable
  - ... the user can choose whether to read in the browser or a dedicated Reading System
Towards bridging online/portable
some topics

1. Generic archive format for the portable state
   - recognized and supported (read/write) by both
     browsers and dedicated reading systems

2. Linking/cross-referencing
   - a generic, registrar-free method for cross-referencing
   - (for the simpler cases at least) no more difficult to employ
     than links on the web

3. Annotations
   - Travel alongside the content, fully agnostic re states and devices
Towards bridging online/portable some topics

4. Metadata
- Unified using HTML-embedded metadata as the vehicle (building on e.g. RDFa and schema.org)
- addressing both discoverability and archival purposes
- ... enabling both publication- and fragment-level expressions

5. Security and privacy model
- For hire: a state-agnostic security model
Presentation Control

- Responsibility for layout control works the same across states and devices
- Publisher has complete control over presentation intent: 100% coverage via CSS
- Personalization: my preferences in an envelope
Content Protection (or not)

- Users will exercise their right to read their content when they want, where they want
- Content protection stands in the way to enable true ubiquity and longevity of e-content
- No apparent universal solution to this problem
Timelines

- DPUB IG charter: September 2015
  - steps after that currently undecided
  - many fundamentals will remain to be designed and implemented
- Through collaboration with W3C, IDPF remains dedicated to providing the portable document format based on the Open Web Platform...

Now (EPUB 3.x) and in the Future (EPUB NG)
Publishing will enhance the web

- Thousands of years of experience in reading, comprehension, communication able to bring to bear on web technology
- Web just beginning to understand…
  - the aesthetics of typography that we know best
  - the user requirements for engaging with long-form portable publications
Thank you for your attention!

Liza Daly, Safari Books Online, liza@safaribooksonline.com
Markus Gylling, IDPF, markus.gylling@gmail.com
Ivan Herman, W3C, ivan@w3.org

http://www.w3.org/2014/Talks/0529-NYC-IH/Talk.pdf