Notes on
W3C Publishing and the Open Web Platform

Summary

"The New Publishing - A W3C Workshop on the Open Web Platform" was held September 16-17, 2013 in Paris. This was an activity primarily of the W3C Digital Publishing Interest Group and was one of several "publishing via the Web" workshops, including a W3C Workshop on Electronic Books and the Open Web Platform in New York this past February. The workshops' purpose was to determine what W3C needed to add to its specifications to enable/encourage total publication development via the Open Web platform. The Paris meeting nominally was concerned with publishing to print.

There were about 68 attendees, including ten W3C personnel. Publishers, publishing organizations and consultants, and universities were largely represented. (List at back of report.)

It was pre-determined that HTML5/CSS was the mechanism. Proponents of XML/XHTML with XSL-FO were assuaged by agreeing that features of that approach needed to be migrated to HTML5/CSS. Observations that examples of viable HTML5 approaches only worked with extensive proprietary extensions indicated that extensions to the basic spec were needed, although there were no specifics nor consideration of how to resolve incompatible proprietary extensions.

The focus on this workshop was nominally publishing to print, and the PWG position was that the CSS should include or reference Print Job Ticket elements, ideally PWG PJT elements, to specify hardcopy production intent. This position was presented in a panel session intended to consider how other standards groups might participate or how their standards could be employed. The organizers thought the PWG position interesting (although leaning toward a JDF job ticket); some W3C CSS types dismissed it as impractical and useless; but the audience seemed uninterested in the idea. One attendee stated that they knew how to get their material printed; she wanted to know how to render the interactive aspects of e-books in printed versions.

Nevertheless, the idea of print job ticket elements in the CSS is on record in our position paper, in the slide presentation which is linked in the W3C workshop agenda, and will be in the workshop minutes and report. Liam Quin (cochair) suggested that there was interest and that the PWG should continue to participate, at least commenting on evolving specifications and possibly joining the W3C. Another W3C person suggested that the CSS group had too narrow a focus, and that the PWG should work though W3C members to introduce this idea at a higher level.

Links:
http://www.w3.org/2012/12/global-publisher/Overview.html - Home page
http://www.w3.org/blog/2013/06/w3c-workshop-industry-consulta/ - blog page
http://www.w3.org/2012/12/global-publisher/agenda.html - Agenda with links to presented slides
http://www.w3.org/2012/12/global-publisher/papers.html - links to submitted papers

Background

This "workshop" was publicized as "an industry consultation event on Publishing", specifically publishing in print as well as e-book formats using the Web and the Web browser interface. The intent is to replace proprietary tools with Web-based applications and with hosted services.

This workshop followed one in February that was specifically concerned with publishing to e-books. (http://www.w3.org/2012/08/electronic-books/). There was remarkably little overlap in attendees (12 out of 68, primarily W3C people), but that might be because of the location difference.

The stated purpose of the workshop was to "hammer out an approach to a solution to Publishing from the web through to printed products." An important question that wasn't clearly addressed was why replace current tools? Various advantages were alluded to during the discussions by proponents, including time and cost efficiency, although it was not clear to what extent the same advantages could be (and perhaps are) provided by non-web services. Indeed, it was pointed out that MS Word is much preferred by authors and editors, and that MS Word has superior change and revision control. However, proponents of an all-web based publishing were very negative about MS Word, referring to it as linear and unstructured. Approaches using MS Word with later conversion to XML, HTML or XTHML were denigrated, with proponents calling for XML early or XML First approaches. The advantages apparently are so obvious that no one thought to state what they were. Of course, XML was also rejected and there was tacit agreement that HTML5/CSS provided the necessary base for Web-based tools, evidentially an outgrowth of the E-Book considerations both at W3C and other standards groups.

Presentations and Concerns

Several of the presentations outlined successful implementation of HTML5 based publishing workflows, but only with proprietary extensions (about which there was some grumbling from CSS folk).

There was a discussion of XML with XLST-FO versus HTML with CCS. Proponents maintained that it had to be HTML with CSS, although it was recognized that many users still use XHTML because they wanted features of XLST-FO which were not available with CSS. One of the identified objectives was to bring up level of CSS to XLST-FO capability.

Although there was much concern over resistance to full adaption of the open web platform by the Publishing industry, only a few expressed the idea that incompatibility with the way authors and editors work and instability/immaturity of the basic specifications were possible reasons for conservative publishers to put off adoption. Indeed, there was some suggestion that the problem was with the authors and editors, who need to be retrained to think non-linearly to allow the system to work well. On the other side, Todd Carpenter (NISO) suggested that W3C was an inappropriate place to be discussing publishing, which prompted W3C people to clarify that they were not getting into tools and workflow (which they were) but were just concerned with the basic WEB capabilities necessary to allow support of these things.

In reference to another point, there was a presentation on, and general agreement that the Browser was the ultimate renderer. It was accepted that once the browser rendered a PDF file, that was tantamount to producing hardcopy. In private discussions, it was agree d that a job ticket was also necessary (JDF being identified), but little consideration was given to conveying the Publishers intent in a CSS. Indeed, one of the chairs (HP) indicated that the CSS did not know (and presumably did not want to know) what fonts were available, and that it was the browser (renderer) that had to make the best effort to address the publisher's rendering (not to say printing) intent.

PWG Presentation

Liam Quin and Karen Myers from W3C were welcoming and repeatedly thanked us for coming. They said that they had regarded the presentation paper as an interesting aspect to be covered.

However, the response to our brief presentation was largely disinterest, although there were several comments.

1. What would be the format content was presented in, HTML or XML? Ans: Either (under assumption that the question dealt with the definition of the product to be printed)
2. Sorry, since the CSS can be changed, whatever is in it can be overridden at printing time and therefore the proposal is useless. Ans: If the task of the print submitter is to produce hardcopy following publisher intent, why would he override the publisher's production instructions?
3. CSS is concerned with formatting. Production instructions should be in the content file, not CSS. Ans (provided by a publisher): No, they use the same content definition file with separate CSS depending upon the form of hardcopy intended…paperback, hardcover, etc.
4. Would the extensions cover other things such as storage? Ans: Perhaps, but the PWG area of expertise and interest is in hardcopy production.

In discussing this response later with Liam Quin, he said that the reference to "communicating publisher production intent", was a red flag that immediately aroused opposition. It was unclear why. He suggested that, if just a few simple examples were given, the concept would have been received better but that the idea of an entire job ticket embedded in the CSS was too drastic. He also showed a definite preference for JDF job tickets.

Other private discussions were, perhaps, more encouraging or at least intended to be. It was observed by Karin that this was perhaps the wrong group to present the idea to ... these were high-end publishers concerned with workflow. My observation was that incorporating production elements in the CSS was very much a Workflow issue…incorporating production instructions in the definition files of a product would allow hardcopy production to proceed more rapidly and with less chance of misunderstanding. She then suggested that it may be more appropriate to introduce the idea at a higher level, and to do so via W3C members.

Workshop Conclusions

The agenda called for a discussion of how to overcome the identified barriers to using the Open Web for publishing though to printed products. It was unclear that barriers had been clearly identified, and comments from the participants indicated some that had not been addressed. What follows are snippets typically relating to problems; there did not seem to be a clearly defined approach.

Problems:

- Metadata: IETF (?) will indicate full list of needs and wants
- Pagination: CSS would like to develop common solution for EBooks and hardcopy. Pagination should also provide for flow within page, as for journals read on line
- Problem with CSS model – normal vs mobile websites – not addressed
- Some publishers using XHTML so they can use XML/XSLT tool kit, rather than HTML and CSS. XSLT probably will vanish from browsers but may hang on
- Ability to change content for different environments (but there was little interest in media queries.)
- Tools for authoring & revision control: Problem of interaction among author /editor/printer – that is a problem in creating spec, and a basic problem to the industry, not a CSS problem. MS Word supports this, and is the leader in change tracking, but switching back and forth (between MS Word and Open Web) is not good. Web tools need change storage.
- Browser as ultimate renderer: Browser must render exactly – no need for PDF reader. Once you have a PDF, the hardcopy is defined.
- How many publishers actually using CSS in workflow: maybe two...four. Possibly more because companies outsource work and may not be aware of tools used. But very little interest in CSS for paged media - no definite outcome from this meeting

Approaches:

- Addressing formatting, pagination, revision tracking with CSS: Dave Cramer will lead effort to identify use cases
- Annotations, reading accessibility and usability, interactivity, , metadata (might be worth separate activity): CSS will address with the Publishing Interest Group (Interest group is member only)
- Web training for publishers: outreach—understand how the parts fit together; Education and outreach- not teaching, but provide overview
- Revision tracking w/formatting: experiment, modify products from Open Office, Delta XML community group

Bill Wagner -26 September, 2013

Organization	Name	Organization	Name
4D CONCEPT	Thomas ROERE	lstituto di Linguistica Computazionale	Marion LAME
4D CONCEPT	Patrice DUBOST	Kozea	Guillaume Ayoub
Adobe Systems, Inc.	Sylvain Galineau	Le Monde diplomatique	Philippe Riviére _
Adobe Systems, Inc.	Alan Stearns	LIP6	Mihnea Tufis
Antidot	Gautier Poupeau	maxdalmas	Massimiliano Dal Mas
Apex	Bill Kasdorf	Mentea	Tony Graham
Apple Inc.	Edward O'Connor	National Information Standards Organization (NISO)	Todd Carpenter
Apple Inc.	Casey Dougherty	Nokia	Frederick Hirsch
Archicol	Emilie Barreau	Nokia	Vlad Stirbu
BISG	Julie Morris	NTT	kawabata taichi
BookSprints.net	Adam Hyde	NTT	Kazutaka YAMAM*OTO*
Byook	Pierre Seriw	0'Reilly Media	Adam Witwer
CEN WS/LT - Oslo and Akershus University College	Tore Hoel	Primento	Mike BOON
Centre Pompidou	Emmanuelle BERMES	Primento	Thibauit LEONARD
Connexions	Philip Schatz	Printer Working Group	Bill Wagner
Corbas	Nic Gibson	Publiwide Ltd	Dmitri Ffilippov
DAISY Consortium	Romain Deltour	Publiwide Ltd	Maurizio Rigamonti
Eden Livres, La martiniere, Le seuil	Vincent Piccolo	Thotm and Soleb Editions	Olivier Cabon
EDltEUR	Graham Bell	UEP	Jirka Kosek
Editions Soleb	Eric Aubourg	University of Bologna	Angelo Di iorio
EDITIS	Nathalie Mosquet	n/a	Katherine Fletcher
EDITIS	Virginie Clayssen	n/a	Gerry Leonidas
Gallimard / Flammarion	Florent Souillot	n/a	Pierre Thierry
Ghent University	Erik Mannens	SNE / French Publishers Association	Flore Piacentino
Ghent University - iMinds - Multimedia Lab	Hajar Ghaem Sigarchian	W3C	Aureiien Chouvet
GRUPO OCEANO	Rafael Gallardo	W3C	Bernard Gidon
Gutenberg Technology	Quentin Valmori	W3C	Bert Bos
Hachette Book Group	Dave Cramer	W3C	Elika Etemad
Hachette Book Group	Phil Madans	W3C	Felix Sasaki
Hachette Livre	Luc Audrain	W3C	Karen Myers
Hachette Livre	Pierre Danet	W3C	Liam Quin
HP	Peter Linss	W3C	Robin Berjon
IDPF & DAISY Consortium	Markus Gylling	W3C	Thierry Michel
		W3C/CWI	Ivan Herman