

PUBLISHING@W3C

Ivan Herman
W3C/CWI, ivan@w3.org

EdTech Conference 2017
16 November, 2017
Madrid, Spain

PUBLISHING@W3C

*Advance the Open Web platform for publishing. Ultimately realizing an ambitious vision for fully Web-native Publications: **publications**—with all their specificities and traditions—**should become first class entities on the Web.***

Broad W3C membership

Some W3C Members that may become, or are already, active in this work



Some history

- ▶ W3C and IDPF has a history of cooperation since 2013.
 - ▶ The “Digital Publishing Interest Group” was established at that time.
- ▶ W3C was also part of the IDPF/IMS “EDUPUB” initiative.
- ▶ W3C and IDPF “merged” in February 2017 to give a completely new life to this cooperation leading to the Publishing@W3C activity.
 - ▶ All the groups that we are talking about have been created a few months ago.

Publishing@W3C Groups

- ▶ *Publishing Business Group*
 - ▶ strategic and business directions
- ▶ *EPUB 3 Community Group*
 - ▶ EPUB 3.1 checker, errata management, new features
 - ▶ “EPUB in Education” (a.k.a. EDUPUB) adaptation to EPUB 3.1
- ▶ *Publishing Working Group*
 - ▶ *towards a W3C Standard on Publishing: Web Publications and Packaged Web Publications*

There is actually more

- ▶ The groups above are the strictly speaking “publishing” groups.
- ▶ However, there are number of issues that
 - ▶ are very relevant to publishing, *but not only*;
 - ▶ whose technical solution is to be worked on by other Working Groups, e.g.:
 - ▶ pagination and general page control — CSS WG;
 - ▶ accessibility issues — ARIA WG and Accessibility Guidelines WG;
 - ▶ offline access — Web Platform WG and Service Workers WG.
- ▶ All Publishing@W3C groups have active relationships with such W3C groups.



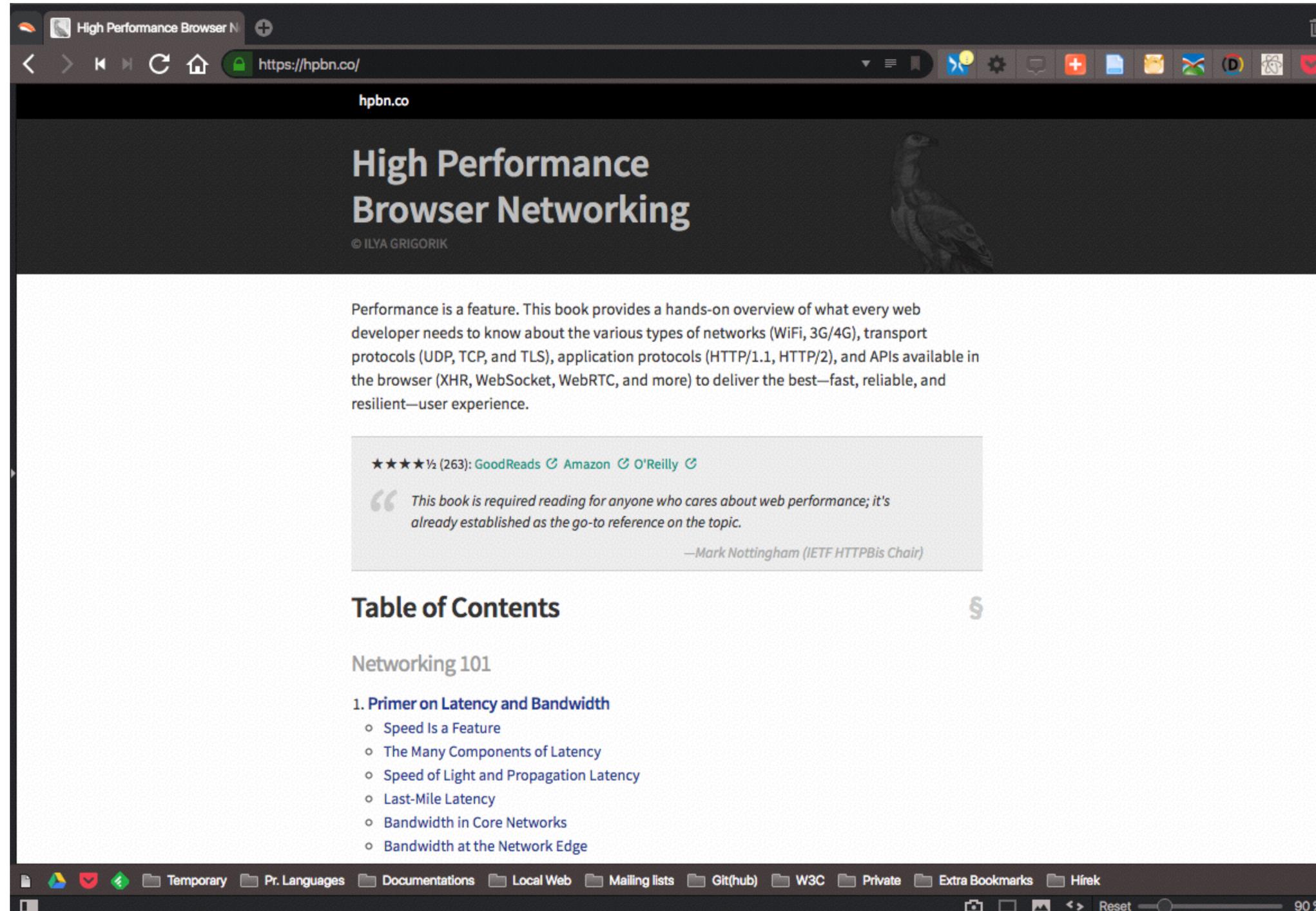
The Publishing Working Group

“Web Publications”: what is the goal?

Publications – with all their specificities and traditions – should become first class entities on the Web.

- ▶ This means:
 - ▶ it should be possible to load the publication content into a browser or a specialized reader, whatever the user prefers;
 - ▶ it should be possible to read the book either offline or online, whatever the circumstances dictate;
 - ▶ contents could be authored regardless of where they are used.

Example: book through a Web browser, visible offline and online



Example: scientific article on the Web

The screenshot shows a web browser displaying a PeerJ article. The browser's address bar shows the URL <https://peerj.com/articles/3902/#p-7>. The article title is "Illegal use of natural resources in federal protected areas of the Brazilian Amazon". The authors listed are Érico E. Kauano^{1,2}, Jose M.C. Silva^{1,3}, and Fernanda Michalski^{1,4,5}. The article was published on October 10, 2017, with a PubMed ID of 29038758. The article is categorized under "Biodiversity", "Conservation Biology", "Ecology", "Environmental Sciences", and "Legal Issues". The abstract section is partially visible, starting with the heading "Background" and the text: "The Brazilian Amazon is the world's largest rainforest regions and plays a key role in biodiversity conservation as well as climate adaptation and mitigation. The government has created a network of protected areas (PAs) to ensure long-term conservation of the region. However, despite the importance of and positive advances in the establishment of PAs, natural resource depletion in the Brazilian Amazon is pervasive." The "Methods" section begins with: "We evaluated a total of 4,243 official law enforcement records generated between 2010 and 2015 to understand the geographical distribution of the illegal use of resources in federal PAs in the Brazilian Amazon. We classified illegal activities into ten categories and used generalized additive models (GAMs) to evaluate the relationship between illegal use of natural resources inside PAs with management type, age of PAs, population density, and". On the right side of the page, there is a "2-year citation median PeerJ articles" badge with the number 5. Below this is a search box for finding colleagues at PeerJ. There are also buttons for "Download" and "Content Alert" (with a "NEW" tag). At the bottom right, there is a "Tools & info" section with links for "Peer Review history", "Citations in Google Scholar", "Ask questions", "Add links", "Visitors" (778), "Views" (1,048), "Downloads" (346), and "Report problem with article". An "Outline" section is also visible with links for "Introduction", "Materials & Methods", "Results", and "Discussion".

These are still not perfect...

- ▶ To achieve this, these publications:
 - ▶ use newest browser technologies (Web Workers, Service Workers, ...);
 - ▶ many aspects are handcrafted in specialized scripts;
 - ▶ or not yet done, e.g., downloading the scholarly article means PDF, i.e., a different format.
- ▶ To achieve a general approach:
 - ▶ lots of details remain to be clarified;
 - ▶ there should be an alignments with the technologies used in Web browsers, by Web developers, ...
- ▶ But: they indicate the direction!

Use Cases

- ▶ Collected by the (now defunct) Digital Publishing Interest Group:
 - ▶ articles (scholarly journals, news agency reports, business publications);
 - ▶ educational materials (text books, exercise collections, etc.);
 - ▶ magazines;
 - ▶ product documentations;
 - ▶ governmental and legal publications;
 - ▶ and, of course, books of any kind.
- ▶ All these use cases raise a number of technical requirements.



An example for an architectural challenge:
many resources under the hood...

Need the concept of “publication” over many resources on the Web

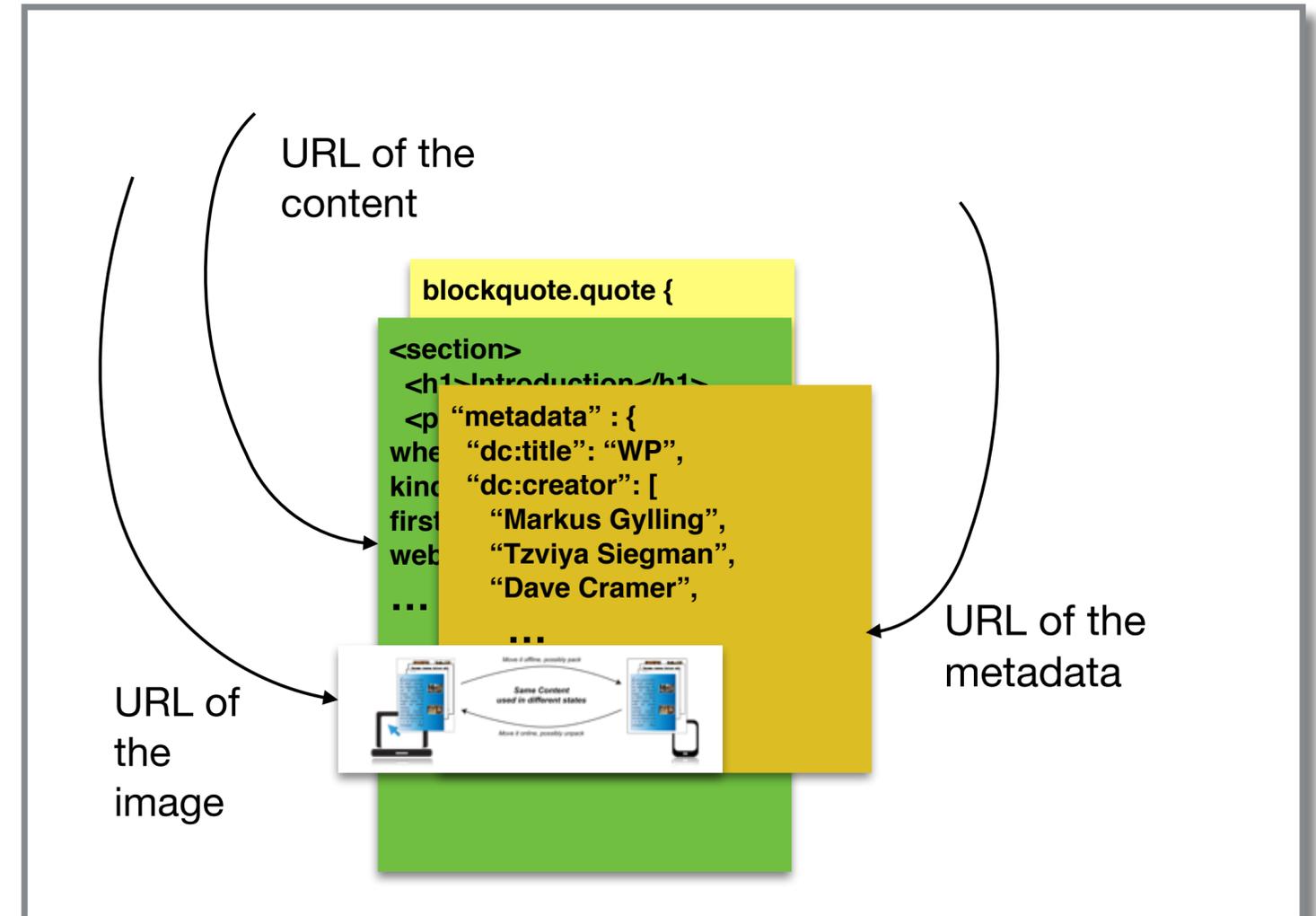
- ▶ A publication may consist of many resources (text, images, videos, etc).
- ▶ On the Web, these are linked together, but they keep their complete “independence”.
- ▶ For publishing the abstract “work” should be a single, conceptual unit on the Web: **a Web Publication (WP)**.

Why do we need the WP concept?

- ▶ The publication has to be checked for integrity as a whole (e.g., in a legal publication).
- ▶ Value counters (counters, section numbering, footnotes, endnotes, etc.) should be consistent across the publication.
- ▶ Any search, text analysis, etc., facility for a book should have access to the full publication.
- ▶ User preferences should apply to the “work” rather than to a single Web page.
- ▶ Publishing and library business models are based on the concept of “work”.

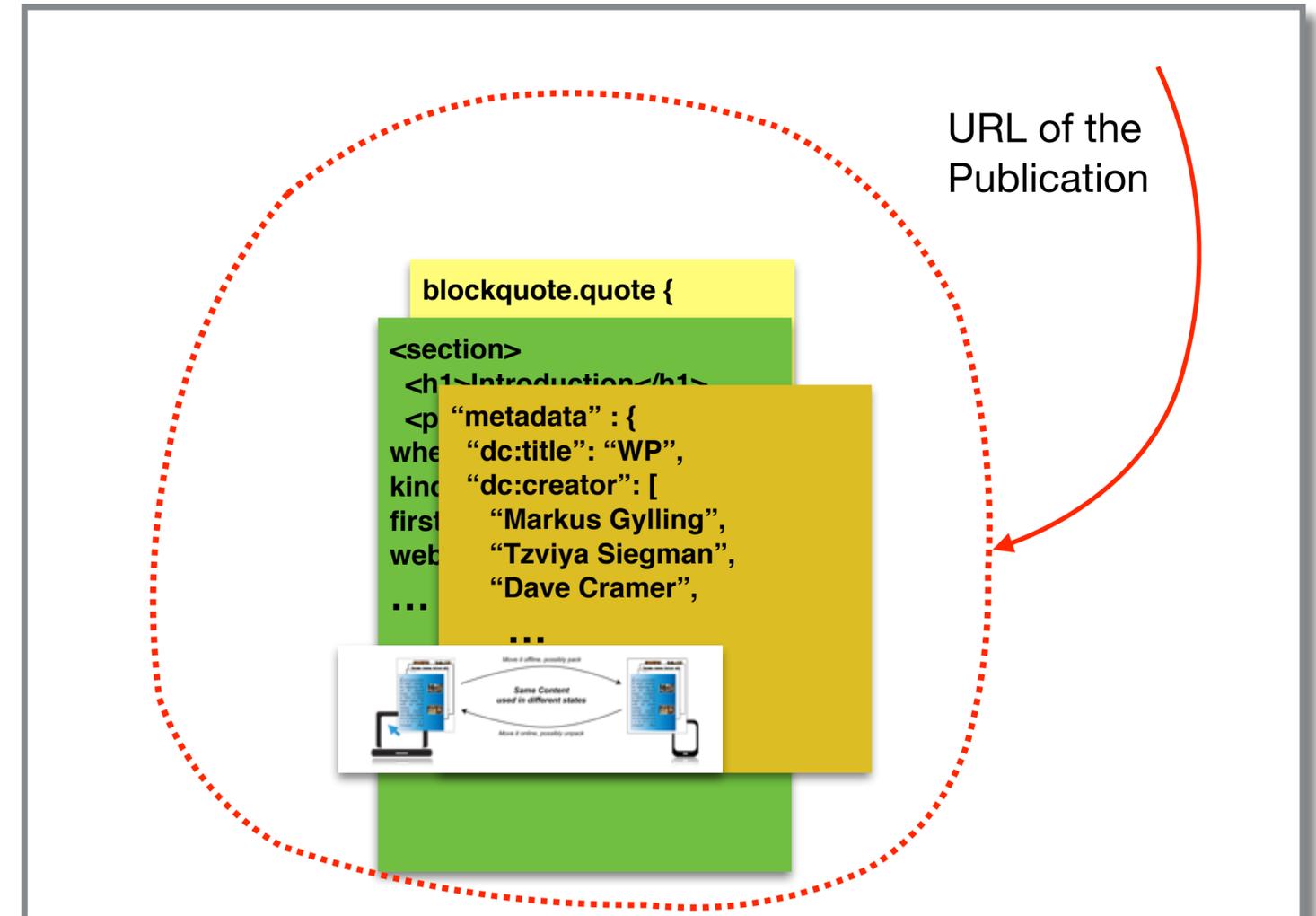
How does that translate to the Web?

- ▶ The current Web builds on the notion of a single resource:
 - ▶ HTML source, metadata, CSS style sheet, etc.;
 - ▶ each has its own URL.
- ▶ Rendering is based on the interoperation of many such resources.



How does that translate to the Web?

- ▶ A WP is a collection of Web resources, considered as a single Web resource
- ▶ there is a unique Web address for the collection;
- ▶ metadata is for the publication *as a whole*.
- ▶ Web Publications are part of the Web!



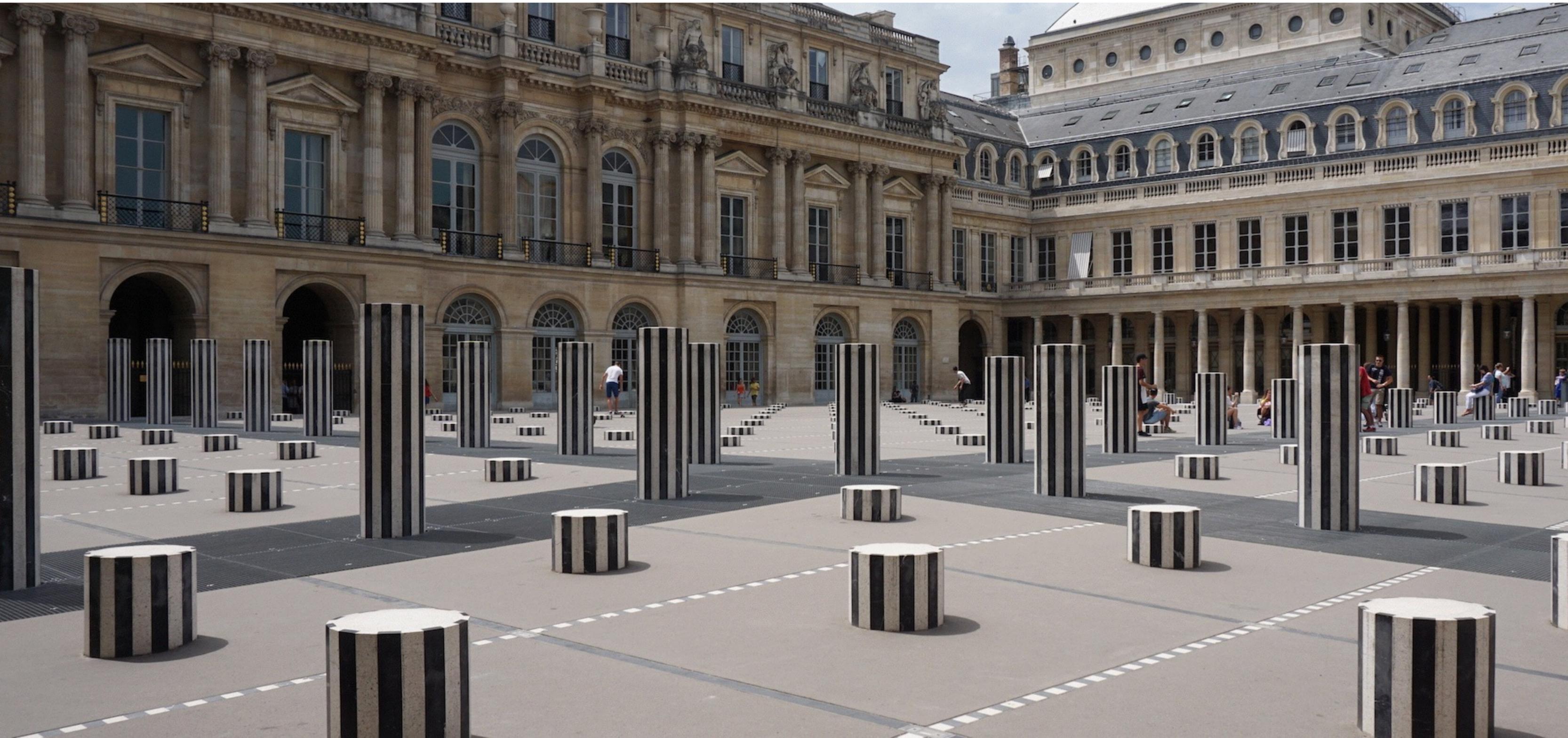
A number of technical issues are being discussed

- ▶ Definition of the “WP InfoSet”, i.e., a collection of information that characterizes a WP
 - ▶ title, constituent Web resources and their reading order, identifiers, default natural language,...
- ▶ Specific security/privacy considerations.
- ▶ Are there specific accessibility issues to consider?
- ▶ What information are necessary for, e.g., unified search?
- ▶ How to achieve offline access to publications?
- ▶ ...

Beyond Web Publications

- ▶ Packaged Web Publication:
 - ▶ the Web Publication may have to be stored and distributed in a separate file (“package”)
 - ▶ what packaging format should be used?
- ▶ EPUB 4:
 - ▶ a *profile* of PWP, with additional requirements (e.g., on accessibility);
 - ▶ achieve *functional* round-tripping with EPUB 3.

Publishing@W3C: combine traditions with modernity...



One more thing before we finish...

This is not how development works at W3C...





...it is more like that!



W3C members move things forward!

Some Useful Links

- ▶ “Home” of Publishing@W3C:
 - ▶ <http://www.w3.org/publishing/>
- ▶ Publishing Business Group:
 - ▶ <https://www.w3.org/publishing/groups/publ-bg/>
- ▶ Publishing Working Group:
 - ▶ <https://www.w3.org/publishing/groups/publ-wg/>
- ▶ EPUB 3 Community Group:
 - ▶ <https://www.w3.org/publishing/groups/epub3-cg/>
- ▶ Web Publication (WG Editors’ draft)
 - ▶ <https://w3c.github.io/wpub/>
- ▶ Web Publications Use Cases and Requirements (Interest Group Note):
 - ▶ <https://www.w3.org/TR/pwp-ucr/>
- ▶ Web Publication (Interest Group Note):
 - ▶ <https://www.w3.org/TR/pwp/>

Slides are available at:

<https://www.w3.org/2017/Talks/Madrid-IH/>

©W3C



Thank you for your attention!