

Great Expectations - The Sequel: Updates from the Digital Publishing Standards Front Lines

Tzviya Siegman, Wiley Karen Myers, W3C 22 March 2018, Toronto

A nod to Charles Dickens – author of great stories and interesting characters



https://en.wikipedia.org/wiki/Great_Expectations



Great Expectations' protagonist Pip speaking with his friend Joe Gargery and wife Georgiana, Pip's eldest sister





HOME

PROGRAM

VIDEO

EXHIBIT HALL

CONNECT

ABOUT



Great Expectations for Digital Publishing with HTML5 and the Open Web Platform

Jeff Jaffe (World Wide Web Consortium (W3C))

4:40pm-5:00pm Wednesday, 02/13/2013

Keynote

Location: Broadway Ballroom

Level:

Publishing is not only a consumer of W3C standards. It is also primary player on the Web and hence a major source of feedback on the technologies in terms of requirements. From production to presentation to distribution, the value chain in the publishing industry has many special requirements that need to be understood to help shape the future of standards in this industry.

Event Partner

Sponsors



Author (R)evolution **Day Sponsor**



Jeff Jaffe

World Wide Web Consortium (W3C)

Dr. Jeff Jaffe is Chief Executive Officer of the World Wide Web Consortium. In this role he works with Director Tim Berners-Lee, staff, and membership, and the public to evolve and communicate the W3C's vision. He is responsible for all of W3C's global operations, for maintaining the interests of all of W3C's stakeholders, and for sustaining a culture of cooperation and transparency, so that W3C continues to be the leading forum for the technical development and stewardship of the Web.



Diamond Sponsors





Platinum Sponsor



W3C Digital Publishing Timeline



Launches W3C Digital Publishing Interest Group



eBooks:

Great Expectations for Web Standards



Hosted by O'REILLY TOC 11-12 Februrary 2013 · NY, NY







First W3C Digital Publishing Workshop February 2013, New York City Hosted by O'Reilly at TOC What Publishing Needs from the Web





eBooks & i18n: Richer Internationalization for eBooks

Second W3C Workshop on eBooks and the Open Web Platform

4 June 2013 · Tokyo, Japan Hosted by Keio University Endorsed by IDPF and support from JEPA members



Workshop Sponsor

W3C Organizational Sponsor

Second W3C Digital Publishing Workshop June 2013

Tokyo, Japan, hosted by Keio University Internationalization Focus

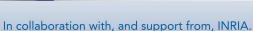




Publishing and the Open Web Platform

Third W3C Digital Publishing Workshop September 2013 Paris, France Open Web Platform, Workflow Focus









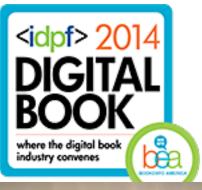


W3C Digital Publishing Timeline

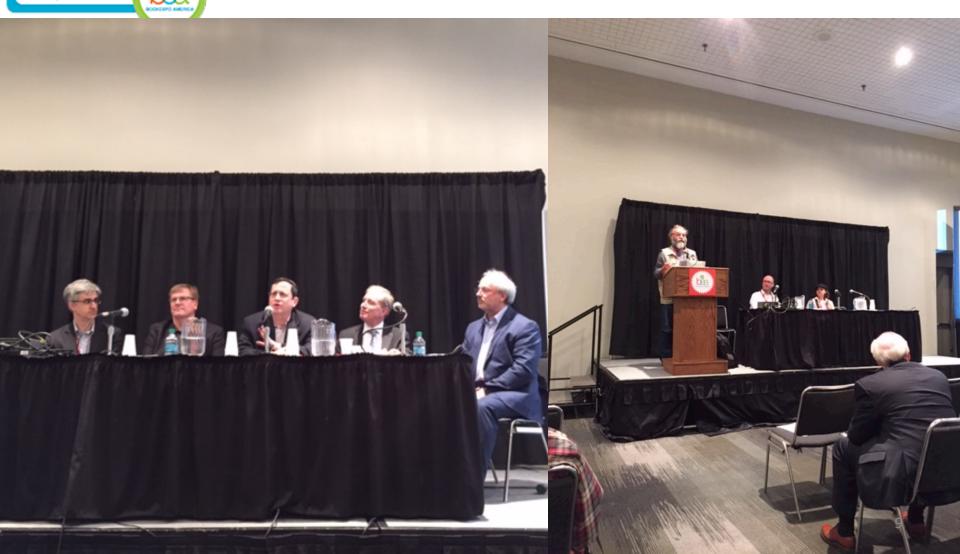


Collaboration with International Digital Publishing Forum

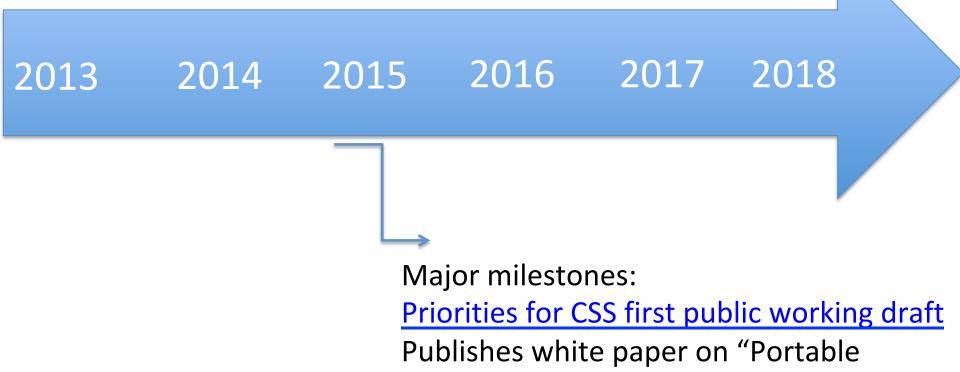
DigPub IG Publishes first draft of <u>Requirements for</u> <u>Latin Text Layout and Pagination</u>



W3C and IDPF Collaborations
Speaking at IDPF DigiCon, NYC
May 2014

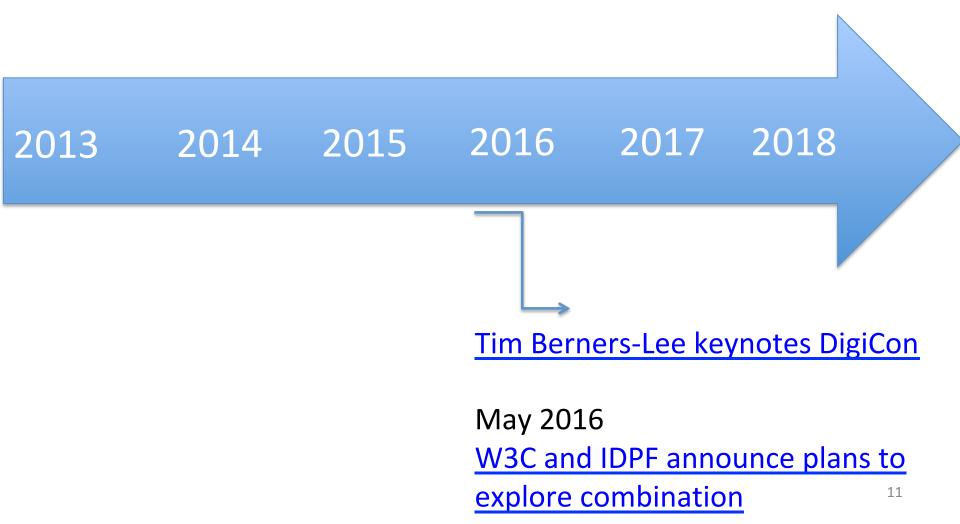


W3C Digital Publishing Timeline



Documents for the Open Web Platform"

W3C Digital Publishing Timeline







FOR IMMEDIATE BELEASE

World Wide Web Consortium (W3C) and International Digital Publishing Forum (IDPF) Explore Plans to Combine

Web Inventor and W3C Director Tim Berners-Lee shares vision of Publishing on the Web in his keynote address at IDPF DigiCon at BEA 2016

Translations I FAQ I W3C Press Release Archive

https://www.w3.org/ — 10 May, Chicago, IL, USA —The World Wide Web Consortium (W3C) and the International Digital Publishing Forum (IDPF) announced today mutual interest in combining their respective organizations to more quickly advance publishing technologies on the Open Web Platform.

Tim Berners-Lee, Web Inventor and W3C Director, and Bill McCoy, IDPF Executive Director unveiled the plans following Berners-Lee's keynote address at IDPF DigiCon at Book Expo America 2016 in Chicago, Illinois, USA.

"We share an exciting vision for W3C and IDPF to fully align the publishing industry and core Web technology. This will create a rich media environment for digital publishing that opens up new possibilities for readers, authors, and publishers," said Berners-Lee.

"Think about educational text books. The book content we know today is becoming highly interactive and accessible with links to videos and images from actual historical events and original research data. This provides greater authenticity and a more engaging learning environment for teachers and students," Berners-Lee commented.

"Whether it's EPUB content or social marketing and reader engagement, Web technologies are foundational to enabling digital content to be created, distributed globally, and viewed on any Web-enabled device as well as dedicated reading apps and eReader devices." McCoy said.

Combining W3C and IDPF Resources Will Achieve Publishing on the Web Vision

Following the announcement, representatives from W3C and IDPF explained that for three years, members from their respective communities have been collaborating to identify how the publishing industry's technology and expertise can improve the Web, and how Web technologies can create greater opportunities for the publishing industry.

"I'm enthusiastic about the prospect of joining forces with W3C. The IDPF's track record of success in developing EPUB standards for the publishing industry will be complemented by W3C's expertise in Web standards that enable accessible rich media," said George Kerscher, President of the IDPF.



Jeff Jaffe, W3C CEO; Tim Berners-Lee, W3C Director; George Kerscher, IDPF President; Bill McCoy, IDPF Executive Director, at DigiCon May 10, 2016

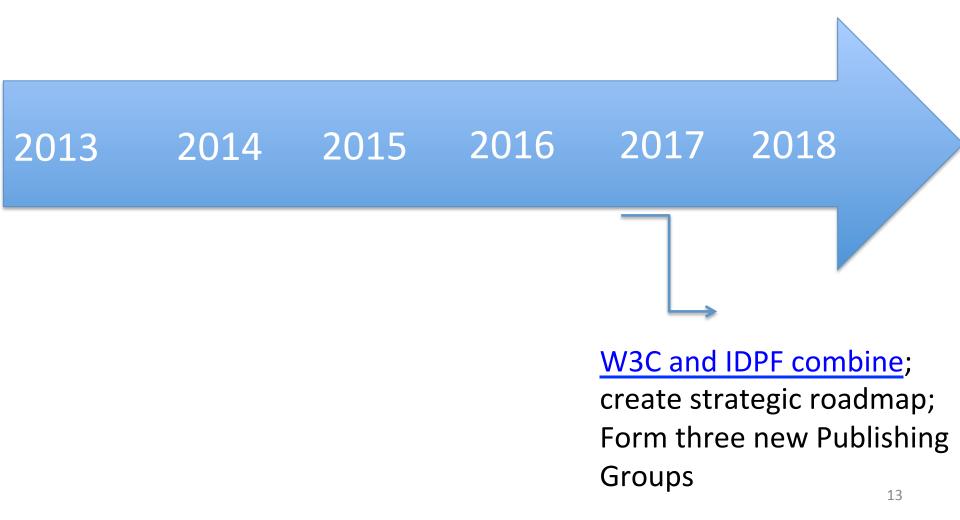
Dr. Jeff Jaffe, W3C CEO, concurred, "With the full participation of the digital publishing community as well as the core web platform developers we accelerate the development of technology standards. Early Web technologies did not meet the rigorous needs of authors and publishers whose curated content is crafted into books, journals and magazines with enhanced styling."

Next Steps

The next steps in the process of combining the two organizations include soliciting comments from the respective memberships of W3C and IDPF and drafting a Memorandum of Understanding. Pending support from both memberships, completion of legal review, and the positive resolution of specific details, the goal is to combine the two organizations by January of 2017. The future evolution of EPUB technical standards would continue at W3C, along with broader work to improve publishing features across the entire Open Web Platform.

About the World Wide Web Consortium

W3C Digital Publishing Timeline





NEWS REVIEWS BESTSELLERS CHILDREN'S AUTHORS ANNOUNCEMED Devices | Copyright | Retailing | Conferences | Content / e-books | Apps

Home > Digital > Content / e-books

W3C and IDPF Officially Combine Organizations

By Calvin Reid | Feb 01, 2017



After seven months of contentious debate over the merger, the International Digital Publishing Forum, developer of the EPUB e-book format standard, and the W3C, the internet standards organization founded by Tim Berners-Lee, have officially combined organizations.

IDPF president George Kerscher, said that the merger of the two organizations will ensure that EPUB will remain a standard relevant to the future of the global online and print community. "By combining our organizations, we not only align our technology roadmaps, but also accelerate the adoption of content that is natively accessible and device-friendly for all types of publishing, whether you are reading on the Web or offline."



Under the merger, the IDPF will be absorbed into the W3C. IDPF members voted overwhelmingly (88% in favor) in support of the merger in November 2016. In addition the IDPF said that "an unprecedented number" of IDPF members who contributed to the development of the EPUB standard have relinquished any rights they have in the standard to ensure that EPUB is an open and royalty free standard. The approval for the merger came despite vigorous objections by Steve Potash, OverDrive CEO and founder of the Open E-Book Forum, the predecessor organization to the IDPF.

W3C Mission:

"Leading the Web to its Full Potential"



Tim Berners-Lee
WEB INVENTOR AND
W3C DIRECTOR

- Founded 1994, now 450+ organization members (corporate, government, other consortia and associations, etc.)
- W3C standards include HTML, CSS, SVG, MathML, XML, XSL-FO, XSLT, ...
- 70+ staff in US (MIT), China (Beihang), France (ERCIM) and Japan (Keio)
- Focus on advancing the overall Open Web Platform, as well as meeting specific industry requirements

Publishing Strategic Roadmap



- Continue to develop and promote EPUB 3 as the packaged accessible interchange and delivery format for digital publications
- Pursue longer-term vision to realize fully Web-native Web
 Publications: online/offline, packaged/distributed, browser/app
- Advance the overall Open Web platform to meet industry needs across segments (trade, education, academic, ...)

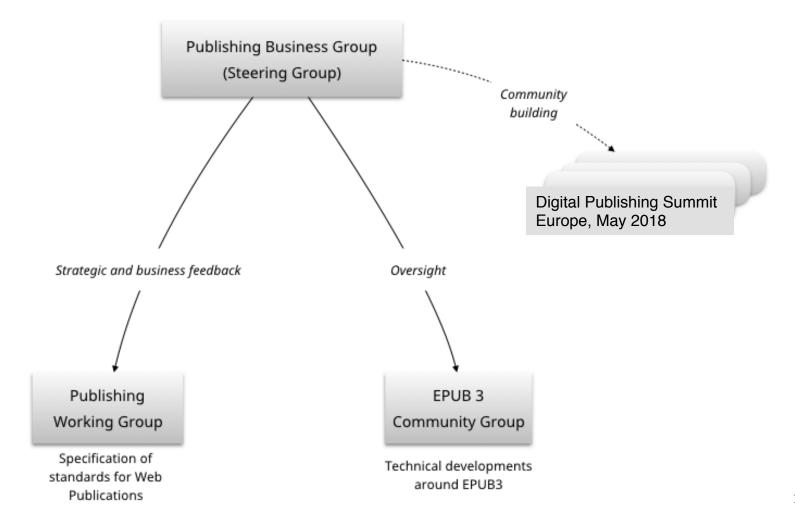
Convergence Vision: Packaged Web Publications

Web Publications, possibly packaged for offline use, share a unified technical architecture and rely on the latest standards of the Open Web Platform.

Publishing requirements are fully and consistently supported throughout the Open Web Platform for packaged publications as well as online Web pages.

Publications are first-class citizens of the Web

How W3C Publishing Activity is Organized



Publishing Business Group – Provides Strategic Direction, Business Focus



Home / Publishing Business Group

PUBLISHING BUSINESS GROUP

The Publishing Business Group fosters ongoing participation by members of the publishing industry and overall publishing ecosystem in the development of the Web for publishing, and serves as a conduit for feedback between the publishing ecosystem and W3C. See the Publishing Business Group Charter for details.

Publishing Business Group Co-Chairs

- Liisa McCloy-Kelley,
 Penguin Random House
- Luc Audrain, Hachette Europe
- Rick Johnson, VitalSourceIngram Content
- W3C Technical staff: Bill McCoy, Ivan Herman

Meets bi-weekly, Tuesdays at 12:00pm EDT.

Accomplishments: Kick-off meeting 13 March 2017 in London, UK; Launch of new Publishing Working Group and EPUB 3 CG

Current focus: Roadmap and task force development

In parallel: Program development for Digital Publishing Summit Europe, May 2018, Berlin, Germany; epubcheck Plan

Participation fee for W3C Business Groups

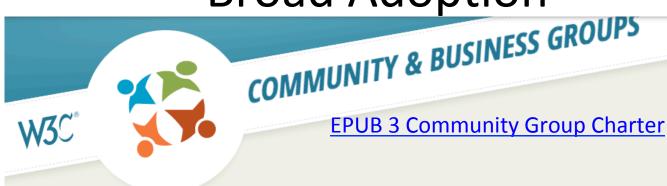
Publishing Business Group Task Forces:

- Accessibility
- EPUB Roadmap
- EPUBCheck fundraising
- Best Practices
- Internal Coordination Liaising with other W3C groups
- External Coordination Liaising with organizations such as BISG, Booknet Canada, DAISY Consortium, EDItEUR, IMS, etc.
- ISO Standardization

Epubcheck

- Validation tool for the entire ecosystem publishers, aggregators, retailers, etc.
- Developers are volunteers
- Future Work needs funding
 - Maintenance Release
 - Support for EPUB 3.2
 - Bug triage
 - Overhaul the test suite to enable easier contribution
- Contact bmccoy@w3.org to contribute funds

EPUB 3 Community Group – Technical Work to Advance EPUB 3 its Broad Adoption



Home / EPUB 3 Community Group

EPUB 3 COMMUNITY GROUP

The EPUB 3 Community Group is a forum for ongoing technical development of EPUB 3 and related extension specifications and ancillary deliverables. The EPUB 3 Community Group charter contains full details.

EPUB 3 Community Group Co-Chairs

- Rachel Comerford,
 Macmillan Learning
- Dave Cramer, Hachette Book Group USA

- EPUB 3.2 spec work
- Best Practices
- Work done on GitHub
- Phone calls as needed
- Free everyone can participate

EPUB 3 Adoption Today

According to VitalSource data for 2017, of the top 100 most used titles in digital textbooks, 77 percent were EPUB 3, and the top 25 were all EPUB files.

"Our more than 22 million users love EPUB files, and the ability to use them online, offline, along with the W3C Web technologies supported in EPUB 3 which are ideal for small screens. Looking at our entire inventory of more than one million titles from 1500 educational publishers, we consistently see 50% to 75% usage every week in EPUB titles."

Rick Johnson, VP Ingram | VitalSource

Publishing Working Group Co-Chairs

- Garth Conboy, Google
- Tzviya Siegman, John Wiley
- W3C Technical Staff: Ivan Herman
- Website: https://www.w3.org/ publishing/groups/publ-wg/

- Charter:
 https://www.w3.org/2017/04/
 publ-wg-charter/
- Make the combination of traditional publishing and the Web complete in terms of accessibility, usability, portability, distribution, archiving, offline access, and reliable cross referencing
- Weekly conference calls
- Work done on GitHub
- Next F2F Meeting: May 2018, Toronto, Canada
- W3C Membership needed

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

- A publication may consist of many resources (text, images, videos, research data, etc.)
- For publishing the abstract "work" should be a single, conceptual unit on the Web: a Web Publication (WP).
 - A WP has its own address, single user agent behavior (e.g., search, value counters), integrity possibly secured for the whole collection as opposed to a single resource, common user preferences, etc.
- "Publication" may mean trade books, textbooks, but also scholarly articles, magazines, governmental and legal publications,...

Requirements of Web Publications

- Addressability: one URL to access the publication, the collection of documents
- Components: what resources are part of the publication?
- **Sequence**: what is the default order of the primary resources?
- Metadata: how to describe the whole rather than the parts
- Personalization: how to allow the reader to adjust the publication's presentation to suit their needs
- Annotations: readers want to highlight, comment, bookmark
- Offline: the state of the network should not control the user experience

W3C Digital Publishing Timeline



Publishing Working Group Publishes First Public Working Drafts (FPWD)

PWG First Public Working Drafts

- W3C Publishing Working Group released 3
 <u>draft specifications</u>: <u>Web Publications</u>,
 <u>Packaged Web Publications</u>, and
 <u>Web Annotations Extension for Web Publications</u>.
- W3C Publishing Working Group
 <u>Technical Summary</u> regular status reports on this wiki

What is new about this?

- Treatment of the publication as a whole many DOMs, but still must be treated as a whole
- Affordances how do make the Web do what we want? What is different from the way the Web works today?
 - Example: search all the files in a WPUB as if it is one document.
- Offline-ability
- Addressability new for publishing

Packaged Web Publications

- All that and a container
- Packaging format is to be determined
- We are considering relying on W3C/IETF work on Web Packaging (in progress)
- The packaging format (at this point) is just sauce

Interested?

- Come join us anyone can follow the <u>email list</u>.
- To participate in the Publishing Working Group, your organization must be a <u>W3C member</u>.
- New Task Forces:
 - Accessibility
 - Affordances
 - Manifest

Advancing the Web Platform for Publishers

- I. User experience
- II. Accessibility
- III. Security / privacy
- IV. Data; Annotations
- V. Immersive Web



I. User Experience

CSS

Advancing styling for high design, high impact content and application experiences

- CSS for print layout
- CSS Grid for responsive
 UX on all devices

Web Fonts (WOFF)

```
raw-option-logo{
 display: block;
 text-indent: -9999
 width: 40px;
 height: 40px;
 background: url()
background-sizes 4
g-option-logo
```

WOFF - Web Fonts



Figure 11. Sample sans-serif fonts

Fonts which are automatically downloaded and used on demand to render a web page, without having to be installed.

The primary purpose of the WOFF2 format is to efficiently package fonts linked to Web documents by means of CSS Open.com <a hr

WOFF File Format 2.0 is New W3C

Recommendation*

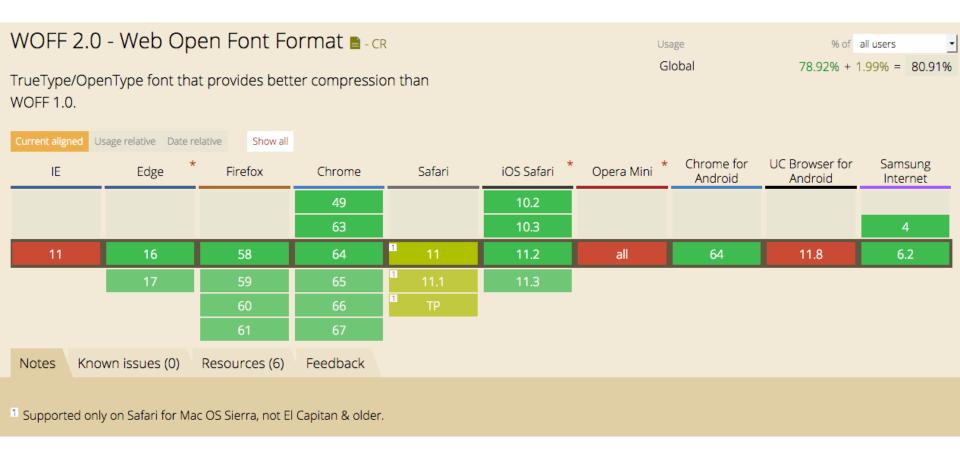
Features

- Significantly improves compression efficiency, lowering use of network bandwidth, and allowing faster compression of font data
- Multicolored OpenType fonts, and OpenType Variable Fonts
- TrueType and OpenType are designed more for ease of access than for minimal file size.

Testing and Use

- Tested against Google fonts corpus, and the Adobe CFF (Type 1) Fonts corpus
- Supported in all major browsers
- 70% of major web sites use
 Web Fonts
- Mobile friendly

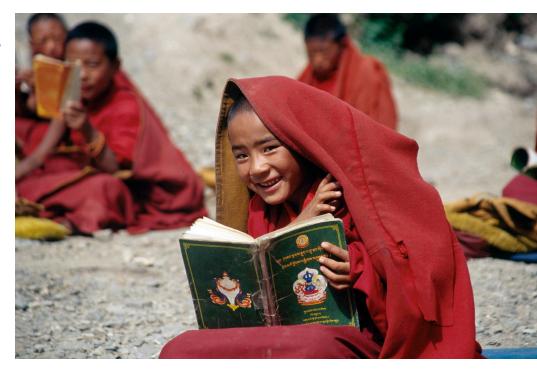
WOFF2 Supported in All Major Browsers



I. User Experience

Internationalization (i18n)

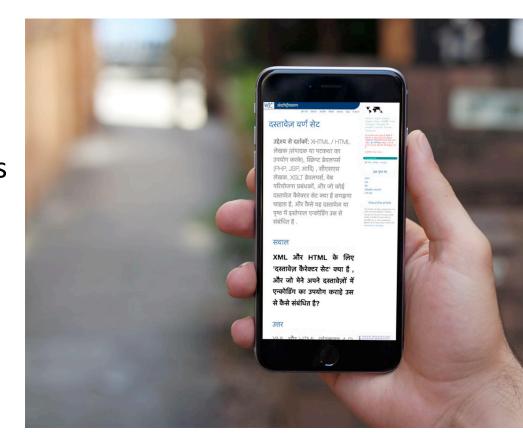
- Horizontal Reviews
 W3C reviews all specifications
 for Internationalization
- Language enablement
 Ensures that local
 requirements for language support on the Web are identified and addressed.
- We welcome your contributions to the <u>language matrix</u> on GitHub.



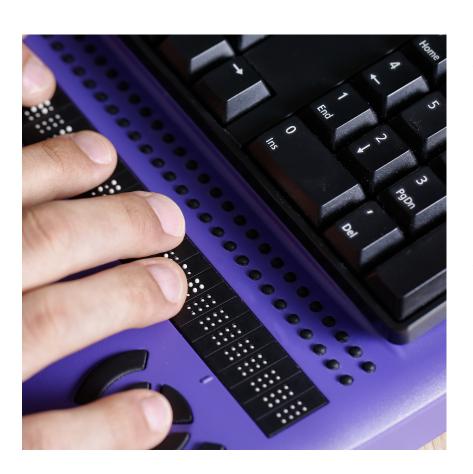
Internationalization (i18n)

Text Layout

Text layout is concerned with rules for line-breaking and justification, local approaches to expressing emphasis or decorating text, localizing counter styles, supporting bidirectional text in markup, initial-letter styling, hyphenation, page layout, and so on.



II. Web Accessibility



Horizontal Reviews

- Accessibility is one of the W3C key areas that get horizontal review of charters and all specifications from experts.
- That means every W3C spec is reviewed for a11y compliance!

How does Web Accessibility affect me?

- Most of us worry about accessible EPUBs (or we should)
- W3C defines Accessibility for the Web. Legislation points to W3C specs, such as WCAG – this is where the action is
- Changes to the specs often mean changes to your workflow. How do you find out about changes?
 Try WAI IG email list
- WCAG applies to EPUB as much as it does to Web sites

Web Accessibility Standards

- The Web Accessibility Initiative (<u>WAI</u>) creates specifications, guidelines, and best practices
- WAI's most well-known works are the <u>ARIA</u> and <u>WCAG</u> specs
 - WCAG 2.1 addresses accessibility of Web content being enhanced with publication-specific semantics as well as low-vision, cognitive and mobile accessibility
 - DPUB-ARIA was co-authored by PWG and ARIA task force
- There are so many more! AAMs, <u>UAAG</u>, <u>ATAG</u>

How can I get involved?

- WAI IG email list is a place for A11y news and questions – announcements about updates to specifications, ask questions; get answers from experts
- WAI Education and Outreach has enough information to get even the total novice started
- Join a group. The best way to keep track of evolving standards is to help write them

III. Security / Privacy

Security

- Security is a major focus of the overall Web architecture; every W3C specification gets horizontal review by security experts
- Web Authentication will soon bring multi-factor authentication to the Web. March 20, 2018 became Candidate Recommendation



III. Security / Privacy

Privacy

- Privacy Activity monitors
 ongoing privacy issues that
 affect the Web, investigates
 potential areas for new
 privacy work
- Content integrity can be tracked reliably through <u>Verifiable Claims</u>

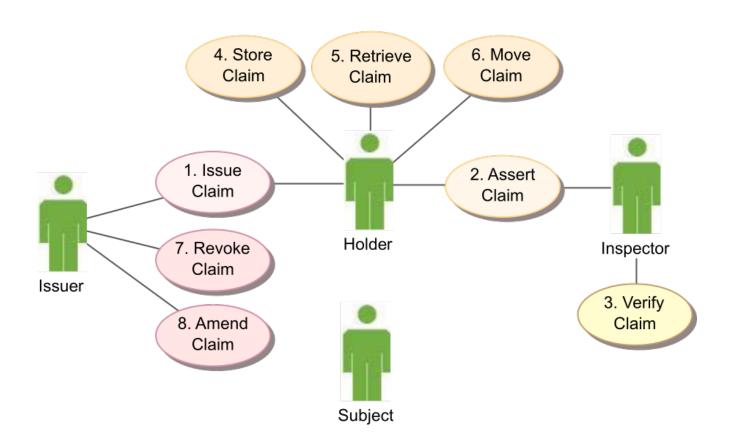


Verifiable Claims

- Credentials are pieces of information about an entity's background
 - driver's licenses are used to assert that we are capable of operating a motor vehicle
 - university degrees can be used to assert level of education
 - government-issued passports enable holders to travel between countries
- Verifiable Claims Data Model provides a mechanism to express credentials on the Web in a way that is cryptographically secure, privacy respecting, and automatically verifiable.

Verifiable Claims Data Model

Verifiable Claims User Tasks



Verifiable Claims Use Cases: Publishing

Educational Test Taking

LMS user needs to identify herself for exam. She is required to provide her credentials to prove identity before the test, and then to allow the system to issue a <u>verifiable claim</u> regarding the results of her test.

Library Subscription

A member of an authorized user community, such as a University Library, would like to access subscription resources. When the user is within the library's physical walls, she can access the materials (from multiple publishers) with her authorized credentials.

When she is on her mobile phone, she can access the same materials remotely. She is required to provide her credentials as a member of the authorized community to access the materials without paying a fee for the content.

IV. Data

Data on the Web

- Web Annotations
- ODRL Permissions,
 Obligations and Expressions
 (POE)
- JSON LD 1.0
- Shapes Constraint Language (SHACLS)
- Data on the Web Best Practices



Web Annotations

To meet the needs of researchers, the project will need to address several key technic considerations, including:

- The diversity of programming environments used in research must b variety of languages and environments are used to perform computations in research and R, to Bash and Matlab, and sometimes multiple languages are combined in a sing The Reproducible Document Exchange Format will be designed with applications like Studio (a popular R Markdown authoring tool) in mind, but ultimately be platform-as enable researchers freedom of choice in their own workflow.
- Researchers should be able to follow their current workflows the rich ecc
 platforms and tools for conducting, storing, documenting and sharing computational
 resource that we intend to build on. Further, some scripts and data are too complex to
 reproducible document. Where possible, we aim to support as many use cases and we
 possible, and we welcome input from the community to help us to understand these
 innovation@elifesciences.org).
- Composing reproducible manuscripts should be easy current authoring too
 Microsoft Word are popular for a reason: What You See Is What You Get (WYSIWYG
 support tools that make authoring and consumption of computationally reproducible
 intuitive as possible for the user. It was with this principle in mind that we recently re
 online journal.

Get involved

We are at the beginning of the process to develop a complete publishing solution to suppression above. We expect this process to take at least 12 more

khinsen

Sep 8

some scripts and data are too complex to sit within a reproducible document

What would be "too complex" for a reproducible document? I can see cases of "too large", in particular for datasets, or for entire virtual machines, but complexity shouldn't be an obstacle.

Hide replies (2)



npscience

Sep 11

Many thanks for your comment.

This statement results from feedback from researchers who could not envisage how their own work could be compiled into this form of document. Complex used here represents an umbrella term, to include: "too large" datasets; computational workflows that are resource-intensive; scripts that are very long (note I am conscious that code quality and conciseness should not present a barrier to sharing).

However, I would say that "are too complex" could be rephrased as "may be too complex", and I would welcome a suggested alternative to "complex" that you feel would be more appropriate.

khinsen



There shouldn't be any limits in principle to a reproducible documents. Workflows that are too resource-intensive may no longer be so in a few years. Better have them in the document, even though few if any can re-run them, than exclude them. Size limits are probably inevitable in a concrete implementation, but should be easily liftable as computing technology progresses. The only reason to exclude infor-

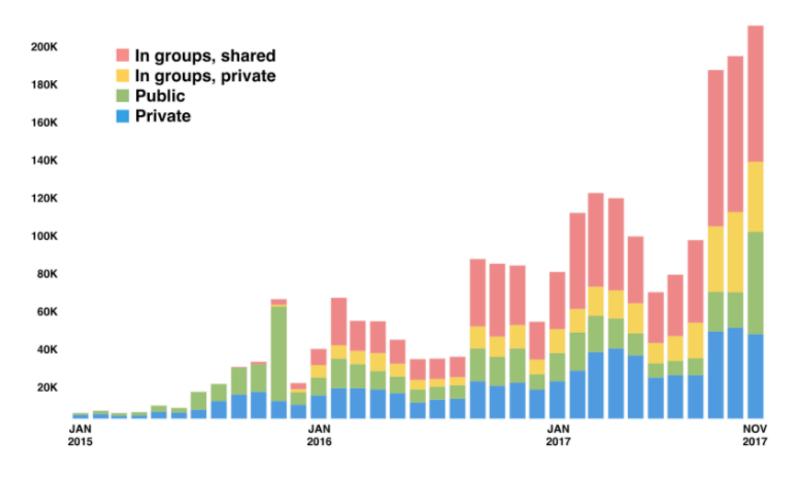
Web Annotations

 Web Annotations allow anyone to annotate any resource on the Web. Annotations can be linked, shared between services, tracked back to their origins, searched and discovered; the vision is for a decentralized and open annotation infrastructure.

Web Annotations became a W3C Recommendation in February 2017

Web Annotations

2.4 Million Annotations and Counting



Permission, Obligations and Expressions (POE)

Open Digital Right Language (ODRL) provides a machine-readable way to describe and express licenses on any Web resource. CC, MIT, W3C, or proprietary licenses can all be expressed in ODRL, and services can be built to provide licensing information for a specific resource.

ODRL became W3C Recommendation in February 2018 under the Permissions, Obligations and Expressions Working Group

JSON LD 1.0

JSON-LD is a lightweight syntax to serialize Linked Data in JSON

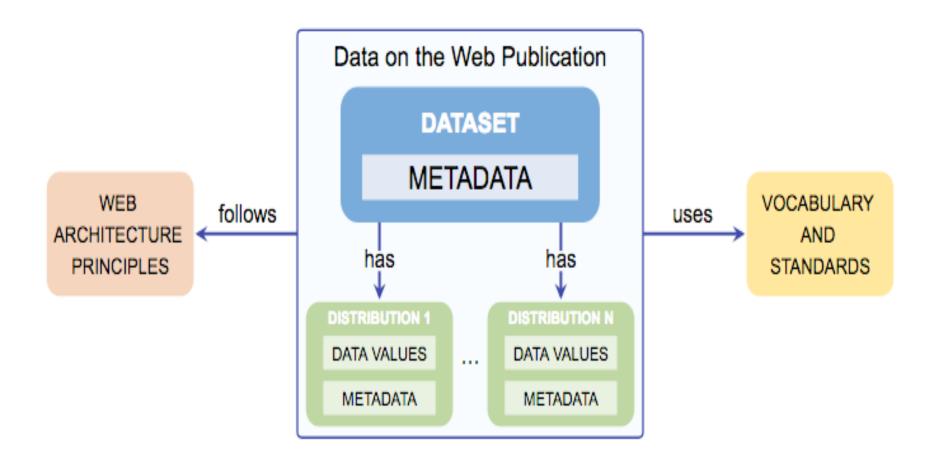
- JSON LD 1.0 became W3C Recommendation in January 2014
- Used, e.g., by Schema.org
- W3C may reopen JSON LD 1.1 to add new features

Data on the Web Best Practices

Data should be discoverable and understandable by humans and machines. Where data is used in some way, whether by the originator of the data or by an external party, such usage should also be discoverable and the efforts of the data publisher recognized.

<u>Data on the Web Best Practices</u> became a W3C Recommendation in January 2017

Data on the Web Best Practices



V. Immersive Web



Virtual Reality

W3C has a WebVR Community Group and expects to charter a Working Group on Virtual Reality in 2018

WebVR will bring new opportunities, particularly for educational publishers

Interesting Cast of Characters

Charles Dickens introduced us to interesting characters like Pip (*Great Expectations*), Quelp (*The Old Curiosity Shop*) and the Artful Dodger (*Oliver Twist*).

Today we've reviewed Publishing@W3C Activities and we have met a few relevant areas for Publishers such as WOFF, POE, i18n, Web Annotations and WebVR.

PUBLISHING@W3C

The next five years...

2018 2019 2020 2021 2022 2023

Publishing Working Group meets F2F in Toronto May 2018

Come meet the interesting people in Publishing@W3C!



Thank you!

Tzviya Siegman, Wiley

Email: tsiegman@wiley.com

Karen Myers, W3C

Email: karen@w3.org