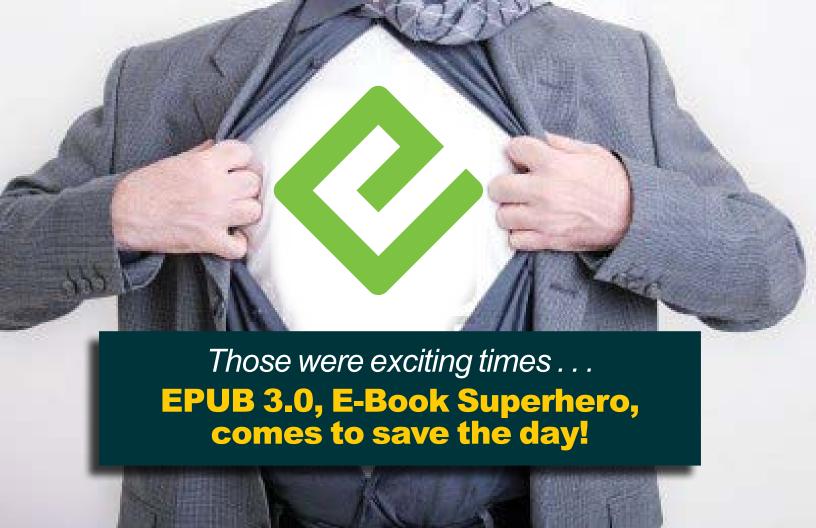


#### **The Continuing Convergence**

Getting from Aspirational to Actual

#### **Bill Kasdorf**

VP and Principal Consultant, Apex CoVantage W3C Publishing WG, PBG, EPUB 3 CG; IPTC; NISO



### **EPUB 3 Was a Huge Advance for the Publishing Ecosystem.**

HTML5!

CSS!

**MathML!** 

JavaScript!

**Audio!** 

Video!

**Lots More!** 





### EPUB 3 Was a Huge Advance for the Publishing Ecosystem.

Specification finalized:

HTML5!

CSS!

**MathML!** 

**JavaScript!** 

Audio!

Video!

**Lots More!** 



### EPUB 3 Was a Huge Advance for the Publishing Ecosystem.

Specification finalized:

HTML5!

CSS!

**MathML!** 

**JavaScript!** 

**Audio!** 

Video!

**Lots More!** 

All features implemented:







#### C'mon, people...

#### Which led to . . .

#### The BISG EPUB Grid / epubtest.org

to show that lots of features of EPUB 3 are implemented in lots of systems, really!

and

#### Readium

to provide an open source implementation that really did implement everything in EPUB 3.



#### Fast forward to today . . .

### PUBLISHING@W3C

Now the work on EPUB and Web Publications has moved to the W3C.



#### Fast forward to today . . .

### PUBLISHING@W3C

Now the work on EPUB and Web Publications has moved to the W3C.

The W3C requires
two independent implementations
of every feature of a spec
before it can become official.



This demonstrates that each feature works in the real world and

two organizations care enough about it to go to the work of implementing it.

The W3C requires
two independent implementations
of every feature of a spec
before it can become official.



#### **Developing a W3C Recommendation**

#### **First Public Working Draft**

"Here's what we're thinking about doing."

#### **Working Drafts**

"Our latest thinking based on feedback and debate."

#### **Candidate Recommendation**

"Okay, we think we've got this figured out now, and how we'll know each feature is implemented."

#### **Proposed Recommendation**

"All features have been implemented or dropped.
Now please vote to advise the Director."

**W3C Recommendation** 

The complexity of the W3C.

Need to align with a multitude of W3C WGs.

#### **Annotations**

Now official and incredibly useful to publishing.

**Rights Expressions** 

Will become essential to publishers.

**Accessibility** 

Web Accessibility is the foundation; being enriched to address the needs of publishers.

**Scores of others** 

The complexity of the W3C.

Need to align with a multitude of W3C WGs.

The need to avoid redundancy.

Don't create new solutions for things that already exist.

#### **Example: Media Overlays**

The EPUB spec for synchronized text and audio, based on a subset of SMIL, now fallen out of favor.

#### **Likely Outcome: Build on Existing Specs**

"OWP appears to have all the tools we need." Work with those various WGs to refine or augment.

The complexity of the W3C.

Need to align with a multitude of W3C WGs.

The need to avoid redundancy.

Don't create new solutions for things that already exist.

The complexity of publications.

A main reason the W3C wants publisher participation.

#### "Publication" ≠ "Document"

A publication is an arbitrarily <u>extensive</u>, <u>complex</u>, & <u>organized</u> <u>collection</u> of resources (web pages, CSS, fonts, images, media, scripts, etc.) with an <u>identity</u>, that can be <u>referenced</u> and <u>archived</u>.

The complexity of the W3C.

The need to avoid redundancy.

Don't create new solutions for things that already exist.

The complexity of publications.

A main reason the W3C wants publisher participation.

The need for implementations.

Features interact, especially in complex publications.

We can't just wait for the browsers.

The browsers want publishers to implement first, to get it right, and to demonstrate demand.

The complexity of the W3C.

Need to align with a multitude of W3C WGs.

The need to avoid redundancy.

Don't create new solutions for things that already exist.

The complexity of publications.

A main reason the W3C wants publisher participation.

The need for implementations.

Features interact, especially in complex publications.

The development of testing.

The publishing supply chain depends on this.

#### So What's Actual and What's Aspirational?

**Really Real Now** 

**Annotations** 

Rights Expressions

Accessibility

Interoperability

Sharing Code

Artificial Intelligence

**Getting Real** 

Resource Access (RA21)

MS Exchange (MECA)

Crossref Co-access

Org Registries

Readium 2

WP/PWP/EPUB4

. . . plus quick updates on upcoming JATS and MathJax developments

#### **The Really Real**



#### **Annotating EPUBs**

Cross-format anchoring: annotate the HTML, PDF, or EPUB of a publication and the annotations appear in the right places in the related other two formats.

\*Hypothesis, a non-profit developing open-source annotation software based on the W3C Open Annotations recommendations, is the leading but not the only annotation implementation.

**Annotating EPUBs** 

**Broadly integrated on systems & platforms** 

Many leading scholarly publishers have integrated Hypothesis, as well as leading platforms like Silverchair, HighWire, Atypon, and Ingenta.

**Annotating EPUBs** 

**Broadly integrated on systems & platforms** 

#### **Publisher Groups**

Default branded and moderated annotation layers on online publications to increase reader engagement, enhance content, and focus attention on the version of record.

**Annotating EPUBs** 

Broadly integrated on systems & platforms

Publisher Groups

**Embedding news videos in annotations** 

Through integration with the Internet Archive's TV News Archive, videos from news broadcasts (or clips) can be embedded in annotations.

**Annotating EPUBs** 

Broadly integrated on systems & platforms
Publisher Groups

**Embedding news videos in annotations** 

#### **OAuth authentication**

Makes login easier & more secure, and integrates with 3rd parties (e.g. social media, ORCID, publishers).

#### **Open Digital Rights Language (ODRL)**

Framework for the machine-readable transmission of rights metadata (NOT access control).

#### **Information Model**

Semantic model for permission, prohibition, and obligation statements.

#### **Vocabulary**

A basic "Core Vocabulary" designed to be augmented.

# Rights Expression

Provides machine-processable expressions for assets, parties, roles, constraints, duties, & actions. Profiles build on these for a specific constituency.

IPTC RightsML

Optimized for news and media industries.
Can be embedded in assets as XML or JSON.

#### **Vocabulary**

A basic "Core Vocabulary" designed to be augmented.

## Rights Expression

Accessibility is the foremost example of the importance of combining EPUB and Web standards.

# Accessibility

#### **Key Current Accessibility Standards**

### EPUB Accessibility 1.0 Accessibility Techniques 1.0

The baseline for accessible publications with guidelines for implementation.

#### WCAG 2.0; 2.1 soon

W3C Web Content Accessibility Guidelines.

#### **WAI-ARIA**

Adds semantics via @role and @aria-\*.

# Accessibility

#### **Accessibility Testing and Certification**

#### **DAISY Ace Checking Tool**

Automated, free, online tool.
Checks for accessibility features, doesn't evaluate.

#### **Benetech Global Certified Accessible**

Comprehensive service: analysis of pubs & workflows and guidance for accreditation and certification.

#### epubtest.org

Evaluates reading systems and assistive technology.

# Accessibility

The future of the publishing ecosystem is all about interoperability.

# Interoperability

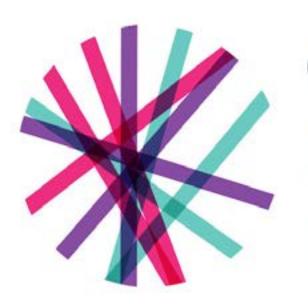
Connecting documents, media, publications, and people: interoperability is fundamental.

Interoperability can only be achieved when based on open standards.

The convergence of publishing and the Web provides a powerful, ubiquitous, & dynamic foundation for the publishing ecosystem.

# Interoperability

#### Toward an interoperable ecosystem



## Collaborative Knowledge Foundation

#### **Coko Foundation**

A community-based, non-profit organization develping modular, open-source workflow components.

#### **PubSweet**

A publication framework with tools for authoring and collaboration, editing and production, workflow management, and administration.

#### INK

An ingestion, conversion & syndication framework.

# Interoperability

#### **Book production workflow**



#### **Editoria**

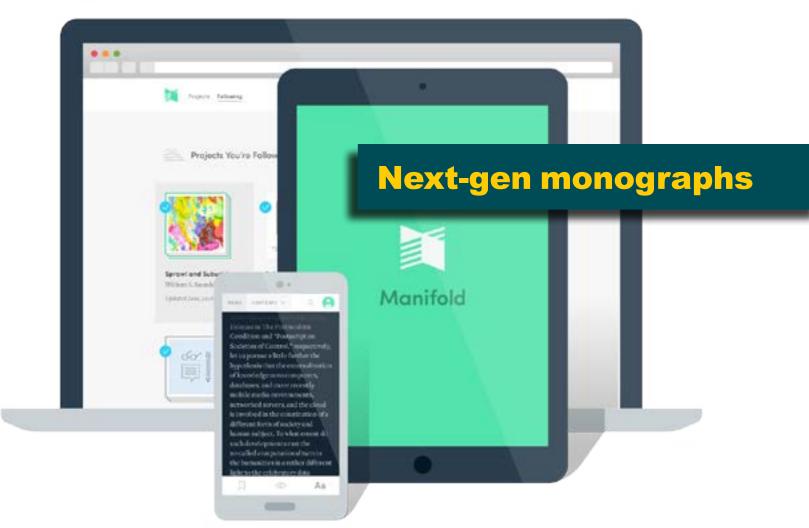
#### Free, open book production infrastructure

Being developed by the Univ. of California Press, California Digital Library, & Coko Foundation.

### Fully integrated, web-based authoring, editing, & typesetting workflow.

Automated conversion from Word to XML/HTML, EPUB- and web-based visual editor, PDF rendering via CSS-based typesetting, & content management.

# Interoperability



#### **Manifold**

#### **Making Monographs "Living Works"**

Being developed by Univ. of Minnesota Press, CUNY Digital Scholarship Lab, & Cast Iron Coding.

Free, open-source, online reading, annotation, discussion, & augmention platform for monographs.

Integrates with print workflows; EPUB- and Web-based. Enrichment with multimedia and other resources.

# Interoperability

**Curation and preservation** 

# fulcrum

#### **Fulcrum**

### Curation, presentation, and preservation of long-form digital scholarship.

Being developed by Univ. of Michigan Library & Press, with Indiana, Minnesota, Northwestern, & Penn State.

### Based on library infrastructure in partnership with the Samvera OS community

Hosted platform for monographs with multimedia, ebook collections, and journals in HSS.

# Interoperability

#### **Sharing scientific code**



### CODE OCEAN

#### **Code Ocean**

Cloud-based computational reproducibility platform.

Researchers can upload code & data and link working code (which gets DOI) with the article.

All computation is done in the cloud, not locally.

Provides reproducibility and transparency.

Code and data can be used for other research too.

# Code Sharing

#### **Getting Real**



#### **RA21: Resource Access for the 21st Century**

Authorizing access via IP address no longer works. People need access anywhere, on any device.

STM & NISO are developing best practices

for providing seamless access to subscribed content with improved UX, security, and text & data mining.

Must work across existing technologies & all players.

Three pilots: one using the WAYF Cloud; one corporate; one based on an identity federation registry.

### Resource Access

#### **MECA: MS Exchange Common Approach**

Re-submission to a new journal is a pain.

MS-mgmt. systems are working on this HighWire, Aries, ScholarOne, eJournal Press, & PLOS

Best practices using existing technologies. Incl. vocab., packaging, metadata, peer review data.

Reference impl. & first live impl. by the end of 2017; needs cross-pub agreements & implementations. Expect to go live in 2018.

# MS Exchange

#### **Crossref Co-access**

A problem that shouldn't be happening but is:

### Books can have multiple DOIs, multiple URLs, multiple depositors.

Publishers and various retailers or aggregators deposit DOIs for the same books or chapters.

Crossref Co-access matches metadata and resolves all DOIs to a Crossref-managed landing page that enables users to pick the link they want.

## Co-Access

#### Find this content at the locations below

#### Clicked Link

Walter de Gruyter GmbH



DE GRUYTER

10.1515/9781400820108

Chapter Title: Adaptation and Natural Selection Author(s): George Christopher Williams Publisher: Princeton University Press DOI: 10.1515/9781400820108 ISBN(s): 9781400820108



**JSTOR** 

10.2307/j.cft7s4g0



Johns Hopkins University Press

10.1353/book.33159

### Co-Access

#### **ORCID OrgID Working Group**

Launched Jan. 2017 based on work in 2016 led by ORCID, Crossref, and DataCite.
Broad and representative membership.
Issued an RFI recently for Registry proposals.

#### **ISNI Organizations Registry**

Announced in August; already > 650,000 org ISNIs. Responded to ORCID RFI and collaborating with other organizations like Jisc, OCLC, CASRAI-UK.

# Org Registries

#### **Readium JS**

#### Readium Cloud Reader

Multi-platform; fetches Web resources from server-side unzipped EPUB (a la "WPs"). Used by NYPL, NYU, Atypon, many others.

Readium Chrome Application Processes zipped EPUBs.

#### **Readium SDK**

A C++ EPUB parser targeting mobile apps.

# Readium

#### In the works: the next generation Readium

#### Readium-2

Current focus: mobile apps.

Native implementations optimized for iOS, Android, & a hybrid for Windows/OSX/Linux: easier to implement, better performance.

iOS & Android SDKs soon as min. viable products.

Readium Desktop to replace Chrome app March 2018.

# Readium

#### **W3C Publishing Working Group**

In first year of a 3-year charter with 4 deliverables:

Web Publications
Packaged Web Publications
EPUB 4
DPUB-ARIA 2.0

FPWDs of WP and PWP expected by end of 2017.

Debate, refinements, and implementations to come!

# Web Publications

### And Two Old Standbys That Keep Getting Better



#### Good standards evolve!

#### **JATS/BITS**

Likely next additions: tagging for CREdit and non-funder contributions.

#### MathJax v3

Early days; just in first public alpha now. Complete overhaul over past 18 months. High level of modularity & performance: expected to be 10x faster!

# Old Standbys

#### Resources

#### **HYPOTHESIS**

Annotating EPUBs: https://web.hypothes.is/blog/epub-annotation/ Publisher Groups: https://web.hypothes.is/blog/publisher-groups/ OAuth: https://web.hypothes.is/blog/leaving-cookies-for-oauth-authentication/ Embedding IA TV News Videos: https://web.hypothes.is/blog/ia-video/

#### **OPEN DIGITAL RIGHTS LANGUAGE**

ODRL Information Model: https://www.w3.org/TR/odrl-model/ ODRL Vocabulary and Expression: https://www.w3.org/TR/odrl-vocab/

#### **ACCESSIBILITY**

EPUB Accessibility 1.0: http://www.idpf.org/epub/a11y/

EPUB Accessibility Techniques: http://www.idpf.org/epub/a11y/techniques/techniques.html WCAG 2.1: https://www.w3.org/TR/WCAG21/

WCAG 2. I. https://www.w3.org/TR/WCAG21/

WAI-ARIA 1.1: https://www.w3.org/TR/wai-aria-1.1/

DAISY Ace Checking Tool: https://inclusivepublishing.org/toolbox/accessibility-checker/ Benetech Certified Accessible: https://benetech.org/our-work/born-accessible/certification/

#### **COLLABORATIVE KNOWLEDGE FOUNDATION**

https://coko.foundation/

#### Resources, cont'd.

**EDITORIA** 

**MANIFOLD** 

**/**/

https://www.fulcrum.org/

**FULCRUM** 

https://editoria.pub/

https://manifold.umn.edu/

**CODE OCEAN** 

https://codeocean.com/

**RESOURCE ACCESS FOR THE 21st CENTURY** 

RA21: https://ra21.org/

**MANUSCRIPT EXCHANGE COMMON APPROACH** 

MECA: https://www.manuscriptexchange.org/

**ORCID OrgID** 

https://orcid.org/content/organization-identifier-working-group

ISNI ORGANIZATION REGISTRY

http://www.isni.org/content/isni-organizations-registry-identifying-organizations-scholarly-supply-chain

**READIUM** 

https://www.edrlab.org/readium/

**W3C PUBLISHING WORKING GROUP** 

https://www.w3.org/publishing/groups/publ-wg/

#### **Thanks!**

#### **Bill Kasdorf**

bkasdorf@apexcovantage.com

+1 734 904 6252

@BillKasdorf

