Pivoting to the Practical

Technology and Standards Focus on Making Things Work

Bill Kasdorf

Principal, Kasdorf & Associates, LLC
Founding Partner, Publishing Technology Partners

THE PIVOT TO THE PRACTICAL

THE PIVOT TO THE PRACTICAL

Focus on what's actually needed.

THE PIVOT TO THE PRACTICAL

Focus on what's actually needed.

Be willing to compromise.

THE PIVOT TO THE PRACTICAL

Focus on what's actually needed.

Be willing to compromise.

Don't over-complicate.

THE PIVOT TO THE PRACTICAL

Focus on what's actually needed.

Be willing to compromise.

Don't over-complicate.

Incubation before standardization.

THE PIVOT TO THE PRACTICAL

Focus on what's actually needed.

Be willing to compromise.

Don't over-complicate.

Incubation before standardization.

Start with a Minimum Viable Product (MVP).

THE PIVOT TO THE PRACTICAL

The bottom line:

Make sure that solutions
—actually work;
—will be implemented by somebody;
—work and play well with others.

Let's start with **the most widely used markup standard in the world**, and one of the biggest:

Let's start with the most widely used
markup standard in the world,
and one of the biggest:

HTML.

Let's start with the most widely used markup standard in the world, and one of the biggest:

HTML.

Did you know that there have been two competing standards for HTML for a long time?

WHAT?!?!?

WHAT?!?!?

Yep, for 15 years, since 2004.

The Web Hypertext Application Technology Working Group (WHATWG) is governed by Apple, Google, Microsoft, and Mozilla: "the browsers."

WHATWG's HTML is a "Living Standard" (LS)

It's continually updated as needs arise; snapshots called "Review Drafts" are issued periodically, typically every 6 months or so.

Features only need implementation commitments.

W3C's HTML is a "Recommendation" (REC)

It's produced by a formal "Process" requiring consensus of the W3C members of each update.

All features must have ≥2 actual implementations.

WHATWG's HTML is a "Living Standard" (LS)

*Agile, responsive, but slipperier and looser.
And driven by the browsers.*

W3C's HTML is a "Recommendation" (REC)

*Rigorous and stable, but slow to adapt.
Driven by a broad range of stakeholders.*

On May 28, 2019,
the W3C and the WHATWG
agreed to collaborate.

Now there will only be <u>one</u> HTML spec.

The work will be done in WHATWG's GitHub, by the WHATWG and the W3C HTML WG.

Annually, W3C selects WHATWG Review Drafts to put through the W3C Recommendation Process.

Features without two implementations will be designated as informative & at risk in the W3C REC.

A practical solution to a longstanding problem.

EPUB has pivoted
to the practical too.

EPUB 3.1 made some important changes.

But it was not backwards-compatible with the widely used EPUB 3.0.1.

It got virtually no adoption.

Existing EPUB 3.0.1s weren't conformant.

Reading systems didn't want to support two specs.

Publishers didn't want to update their EPUB 3.0.1s.

There are thousands of EPUB 3.0.1s out there!

EPUB 3.1 made some important changes.

But it was not backwards-compatible with the widely used EPUB 3.0.1.

It got virtually no adoption.

The mistake:

standardization before implementation.

There are thousands of EPUB 3.0.1s out there!

The W3C published EPUB 3.2 on May 8, 2019.

Strongly *backwards compatible* with EPUB 3.0.1.

Kept useful improvements from EPUB 3.1.

No longer specifies specific versions of *HTML, CSS, and SVG*. Now *stays in synch* with those specs.

Recommends *accessibility conformance*.

Spec documentation reorganized and improved.

Almost all EPUB 3.0.1s conform to EPUB 3.2.

Why am I supposed to care?
We mainly publish journals.

Millions of journal articles will be published as EPUBs by the end of the year.

That means they're much more accessible than PDFs.

But accessibility isn't the only reason to move from PDF to EPUB as the main format.

They're just plain better.

"Improved accessibility benefits everyone.

"While users of AT may realize the greatest benefit, the responsiveness (content reflowing based on the device viewport size), semantics (schema.org and other vocabularies), interactivity (embedded audio, video, and other media), and reproducibility (embedded data and executable code) represent a significant advance in both preserving and disseminating the scholarly record."

—Joshua Pyle, Atypon

The Web Publishing work in the W3C has also pivoted to the practical.

Web Publications spec published as a WG Note along with a *WP Use Cases and Requirements* Note in Aug.

Lightweight Packaging Format is also a WG Note, referencing the *Publication Manifest,* on Rec track as Candidate Recommendation (CR) to facilitate *implementations of profiles* of Web Publications.

First profile, Audiobooks, is on the Rec track as a CR.

The Web Publishing work in the W3C has also pivoted to the practical.

Main WP Rec track work is paused for a reason.

Make sure that solutions
—actually work;
—will be implemented by somebody;
—work and play well with others.

Implementation before standardization.

W3C Community Groups have become key to incubating implementations.

The EPUB 3 CG developed EPUB 3.2.

There is a new Publishing CG to facilitate implementations of Web Publications.

The MathML Refresh CG is working on better aligning MathML with the Web Platform and CSS.

A key to CGs: anybody can join, no dues.

THREE *(obvious!)* STRATEGIES FOR PIVOTING TO THE PRACTICAL

1. Focus on what's needed.

2. Create tools.

3. Update based on current reality.

THREE *(interesting!)* **EXAMPLES OF FOCUSING ON WHAT'S NEEDED:**

1. **RA21:** *what's needed is a good UX.*

2. **LMT:** *what's needed is precision.*

3. **ROR:** *what's needed are IDs for research orgs.*

[IOI, Invest in Open Infrastructure, is another one; since I discussed them last year I won't repeat that.]

Focus on what's needed.

RA 21

has led to

THE COALITION FOR SEAMLESS ACCESS

RA21's work resulted in the recommended practices that it set out to create: NISO RP-27-2019.

A key standard, SAML ("Shibboleth"), effectively addresses the issues around federated authentication.

But poor user experience has led to low adoption.

The Coalition for Seamless Access was created to develop a clear, consistent, trusted process.

Focus on what's needed.

THE COALITION FOR SEAMLESS ACCESS

Collab. between GÉANT, Internet2, NISO, ORCID, STM.

SeamlessAccess.org is a service that leverages existing single-sign-on infrastructure.

Based on RA21 Recommended Practices.

Provides a consistent, clear UX between systems.

Strictly protects personal data and privacy.

Focus on what's needed.

THE COALITION FOR SEAMLESS ACCESS

And yes, this is real:

SpringerNature is the first publisher to deploy this in production.

Provides a consistent, clear UX between systems.

Strictly protects personal data and privacy.

Focus on what's needed.

LMT, the Language Metadata Table

MESA — Media & Entertainment Services Alliance

LMT, THE LANGUAGE METADATA TABLE

Makes it easier to apply precise language codes.

Developed by HBO, now maintained by MESA and virtually all the big media companies.

Based on IETF BP 47 (ISO 639 lang. codes, ISO 3166 country codes, & UN M. 49: UN Territory stds.), and it's a WWW standard supported by the W3C.

LMT enables expressing language, context, dialect.

Focus on what's needed.

LMT, THE LANGUAGE METADATA TABLE

Language groupings keep subgroups together.

- *Greek*: to account for ancient vs modern.
- *English*: British, Canadian, Australian, American, etc.
- *Spanish*: Latin American vs European, Mexican vs Argentinian

Focus on what's needed.

LMT, THE LANGUAGE METADATA TABLE

*Full syntax enables *great precision*:*
language-script-region-variant-extension-privateuse.

*For example, *mn-Cyrl-MN* indicates Mongolian written in Cyrillic as used in Mongolia.*

Focus on what's needed.

LMT, THE LANGUAGE METADATA TABLE

Includes *endonyms* (language name in the country's language): *e.g., Français.*

Includes *exonyms* (language name as spoken in other countries): *e.g., Französisch.*

Discovery & distribution are global!

Focus on what's needed.

ROR

The Research Organization Registry

There are several org IDs (e.g., ISNI, Ringgold, GRID), created for different purposes; some proprietary.

Crossref DOI is a content identifier.

ORCID is a contributor identifier.

ROR ID is a <u>research organization</u> identifier.

ROR, the Research Organization Registry, maintains ROR IDs and associated metadata.

Focus on what's needed.

ROR's aim is to be
"… a *community-led* project to develop an
open, *sustainable*, *usable*, and *unique identifier*
for *every research organization* in the world."

ROR ID is a research organization identifier.

ROR, the Research Organization Registry,
maintains ROR IDs and associated metadata.

Focus on what's needed.

THE RESEARCH ORGANIZATION REGISTRY

Led by California Digital Library, Crossref, DataCite, and Digital Science, with 40+ advisors.

ROR MVR (Minimum Viable Registry) launched in January 2019 with ROR IDs for 91,000+ organizations.

Initially based on and syncs to GRID database; maps to GRID, ISNI, Crossref Funder ID, and Wikidata.

ROR IDs and metadata are open (CC0 license).

Focus on what's needed.

THE RESEARCH ORGANIZATION REGISTRY

*ROR scope is the "affiliation use case":
"contributors, contributions, research sponsors,
publishers, and employers."*

*Research Organizations are defined as
"any organization that conducts, produces,
manages, or touches research."*

Focus on what's needed.

THREE *(useful!)* **EXAMPLES OF CREATING TOOLS:**

1. **Ace by Daisy:** *automated accessibility checker.*

2. **KFG's PubPub:** *a sandbox for innovation.*

3. **ARTiFACTS:** *establishing research provenance.*

Create tools.

Ace by DAISY

ACE BY DAISY

Automated checking of EPUB accessibility (i.e., conformance to *EPUB Accessibility 1.0* spec).

Free and *open source*.

Available as a *command line tool* to be integrated into workflows and as a *desktop app*.

Already *widely used* by EPUB *production services* as well as *publishers* and *individuals*.

Create tools.

ACE BY DAISY

Provides summary + extremely detailed report:
—Groups issues by category;
—Shows where each instance occurs;
—Links to Knowledge Base to provide guidance;
—Indicates where human attention is needed.

Very fast, easy to use and to understand.

Create tools.

The Knowledge Futures Group's PubPub

PubPub

PUBPUB

Free, open source hosting platform.

Create tools.

PUBPUB

Free, open source hosting platform.

Free, open source *authoring platform*.

Create tools.

PUBPUB

Free, open source hosting platform.

Free, open source authoring platform.

Free, open source *collaboration platform*.

Create tools.

PUBPUB

Free, open source hosting platform.

Free, open source authoring platform.

Free, open source collaboration platform.

Free, open source publishing workflow platform.

Create tools.

PUBPUB

Free, open source hosting platform.

Free, open source authoring platform.

Free, open source collaboration platform.

Free, open source publishing workflow platform.

Free, open source multimedia publishing platform.

Create tools.

PUBPUB

Free, open source hosting platform.

Free, open source authoring platform.

Free, open source collaboration platform.

Free, open source publishing workflow platform.

Free, open source multimedia publishing platform.

It's all of the above.

Create tools.

It's a sandbox.

PUBPUB

Free to start a community—"just start using it."

Provides a wealth of functionality: peer review mgmt., annotation, citation mgmt., visualizations, & more.

Used by >650 communities: journals, books, conferences, interactive/multimedia projects, interdisciplinary projects, collaborative review.

Facilitates experimentation & innovation.

Create tools.

Establishing Provenance for Research Outputs

ARTiFACTS

Preserves "a valid and immutable chain of records" of all stages of the research lifecycle, all research outputs, artifacts, transactions, citations, etc.

Publication just preserves the end of that lifecycle.

ARTiFACTS is a blockchain-based platform that establishes provenance from the beginning.

Captures & links info. from ideation to dissemination.

Create tools.

ARTiFACTS

Recently integrated with ORCID.

Now all researchers can register research outputs and artifacts using their ORCID credentials, establishing "proof of existence and ownership" of their work and facilitating citation.

Enables and promotes secure sharing.

Create tools.

TWO *(timely!)* **EXAMPLES OF UPDATING BASED ON CURRENT REALITY:**

1. **Readium:** *because of <u>ecosystem changes</u>.*

2. **MathML:** *because of the <u>dominance of browsers</u>.*

Update responsively.

Free, Open Source
EPUB Reading System Software

READIUM

The diversity of the EPUB ecosystem made "one master version works everywhere" impossible. And Google dropped support of Chrome apps.

Now Readium is "a complete set of reading toolkits" for mobile, desktop, and web applications.

EDRLab has developed the Thorium Reader, the first application based on Readium Desktop.

Update responsively.

READIUM

All toolkits support *EPUB 3.2* and *OPDS,* the Open Publication Distribution System.

The Android toolkit supports *audiobooks.*

Support for *Divina (Digital Visual Narratives)*— digital comics and manga—in development.

Internal format contributes to *Web Publications.*

Update responsively.

W3C MathML Refresh
Community Group

MATHML REFRESH

A *W3C CG* formed Jan. 2019 to improve MathML:
- Better *align with the current web* environment;
- *Ease the burden on browser* implementations;
- Increase support for *assistive technology*.

Working on two specs:
- *MathML Core*—to be implemented in *browsers*;
- *MathML 4.0*—keep *compatibility with MathML 3.0*, including *deprecated elements* and *unused attributes*.

Update responsively.

MATHML REFRESH

MathML in Chromium *(basis for Chrome & Edge)*
Sloan-funded by NISO; *work being done by* Igalia.
Based on MathML Core; *fully* W3C conformant.
Active contributor to MathML Refresh CG.
Currently well into the implementation phase.

Update responsively.

THE PIVOT TO THE PRACTICAL

The bottom line:

Make sure that solutions
—actually work;
—will be implemented by somebody;
—work and play well with others.

Resources

WHATWG
https://whatwg.org/

W3C HTML 5.2
https://www.w3.org/TR/html52/

EPUB 3.2
https://www.w3.org/publishing/epub/

WEB PUBLICATIONS
https://www.w3.org/TR/wpub/

WEB PUBLICATIONS USE CASES AND REQUIREMENTS
https://www.w3.org/TR/pwp-ucr/

PUBLICATION MANIFEST
https://www.w3.org/TR/pub-manifest/

LPF, LIGHTWEIGHT PACKAGING FORMAT
https://www.w3.org/TR/lpf

AUDIOBOOK PROFILE
https://www.w3.org/TR/audiobooks/

Resources

COALITION FOR SEAMLESS ACCESS
https://seamlessaccess.org/

LMT, LANGUAGE METADATA TABLE
https://www.mesalliance.org/language-metadata-table

ROR, RESEARCH ORGANIZATION REGISTRY
https://ror.org/

ACE BY DAISY ACCESSIBILITY CHECKER
https://inclusivepublishing.org/toolbox/accessibility-checker/

PUBPUB
https://www.pubpub.org/

ARTiFACTS
https://artifacts.ai/

READIUM
https://readium.org/

MATHML REFRESH / MATHML IN CHROMIUM
https://www.niso.org/niso-io/2019/07/update-mathml-chromium-project

Thanks!

Bill Kasdorf

kasdorf.bill@gmail.com

pubtechpartners.com

+1 734-904-6252

@BillKasdorf