Collaboration & Convergence

How Today's Technology Standards are Working Together to Make Things Work

Bill Kasdorf

VP and Principal Consultant, Apex Content Solutions

What the standards landscape has too often seemed like.

Why are so many standards silos?

Because they're created within specific domains or specific communities of interest to do specific things.

APEX
Content Solutions

Why are so many standards silos?

Because they're created
within specific domains or
specific communities of interest
to do specific things.

This is not a bad thing.

It's usually a pretty good thing . . .

. . . to do <u>that</u> stuff

. . . for <u>those</u> folks.

APEX
Content Solutions

DOI
NLMXML
MARC
PDF

Like these, which we've all come to know and love.

What's different now?

Now everything seems to be connected to everything.

Disconnects are dysfunctional.

APEX
Content Solutions

What's different now?

Now everything seems to be connected to everything.

Disconnects are dysfunctional.

People expect

INTEROPERABILITY

RELIABILITY

PORTABILITY

ADAPTABILITY

APEX
Content Solutions

INTEROPERABILITY

Enabling systems and processes to interoperate smoothly and predictably removes friction, resulting in efficiency, economy, and flexibility.
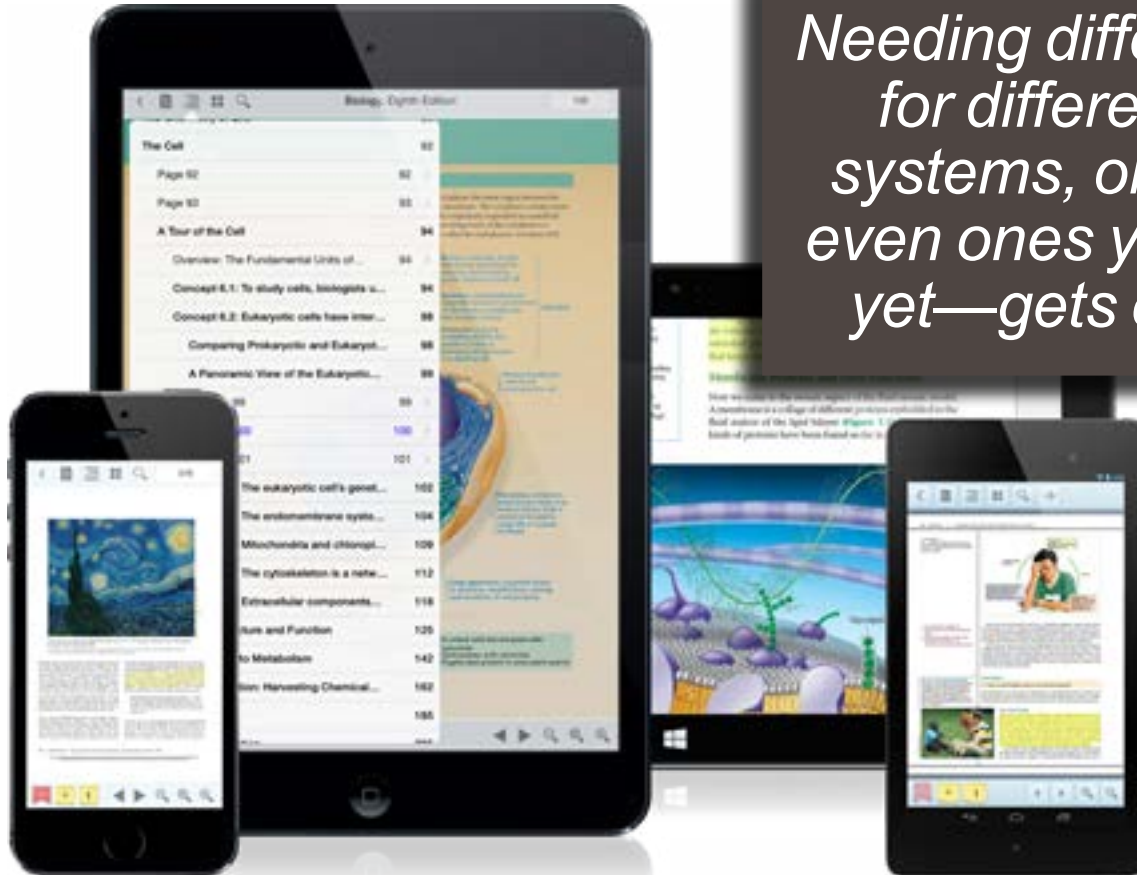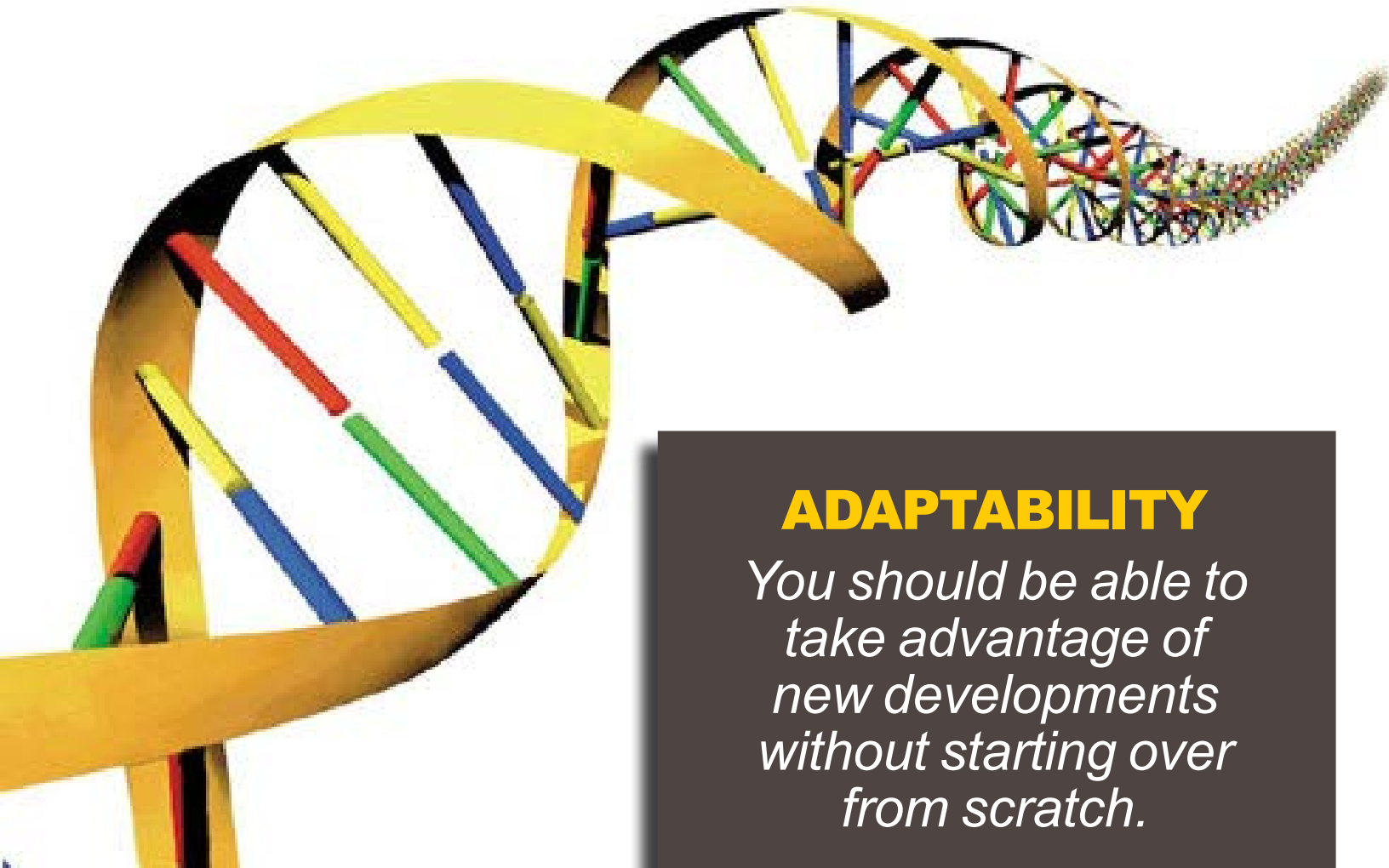
RELIABILITY

*Things need to "just work"
when you move
or access content
between devices and
systems.*

PORTABILITY

Needing different versions for different devices, systems, or platforms— even ones you don't have yet—gets old real fast.

ADAPTABILITY
You should be able to take advantage of new developments without starting over from scratch.

Standards provide

the **oil that removes friction**,

the **"common language"**
that lets systems understand each other,

and the **broad base of support** and use
that keeps them up to date.

I don't want to change what I'm doing.

Isn't that what transforms are for?

Can't I just transform what I've got to what somebody else needs?

Well, sure.
But there's a catch. Several, actually.

Well, sure.
But there's a catch. Several, actually.

Things don't always map cleanly.

APEX
Content Solutions

Well, sure.
But there's a catch. Several, actually.

Things don't always map cleanly.

Transforms are often not reversible.

APEX
Content Solutions

Well, sure.
But there's a catch. Several, actually.

Things don't always map cleanly.

Transforms are often not reversible.

**It's easy to miss things
in unexpected contexts.**

APEX
Content Solutions

Well, sure.
But there's a catch. Several, actually.

Things don't always map cleanly.

Transforms are often not reversible.

It's easy to miss things
in unexpected contexts.

People mess up.

Well, sure.
But there's a catch. Several, actually.

Things don't always map cleanly.

Transforms are often not reversible.

It's easy to miss things
in unexpected contexts.

People mess up.

Transforms are essential.

**But they work best
when input and output
are as similar as possible.**

APEX
Content Solutions

Standards built on standards—

open, non-proprietary, interoperable standards

—are becoming a fundamental requirement.

APEX
Content Solutions

WEB STANDARDS

THE FOUNDATION OF TODAY'S DIGITAL ECOSYSTEM

The Web is 25 years old, and the W3C is 20.

W3C®

Vint Cerf

Sir Tim

A Few of the Key Web Standards STM Uses

XML
A meta-language for defining markup schemes

HTML
The markup language of the Web, browsers, etc.

CSS
The stylesheet language for rendering HTML

MathML
The language for encoding math as XML

SVG
The format for encoding images as XML

A Few of the Key Web Standards STM Uses

XML
A meta-language for defining markup schemes

**JATS/BITS (née NLM)
is an XML markup scheme/tagset.**

So are DocBook, TEI, and DITA. And WileyML.

So is XHTML.

Yep. It's real XML. It's HTML expressed as XML.

SVG
The format for encoding images as XML

XML.

Models Made with XML.

A Few of the Key Web Standards STM Uses

XML
A meta-language for defining markup schemes

HTML
The markup language of the Web, browsers, etc.

The long-awaited **HTML5** recommendation was finally official as of October 28, 2014.

But it has been under development for years and is already in use in today's **browsers, apps, platforms—and EPUB 3**.

HTML5 is the structural markup

. . . and the behaviors expected from browsers based on that markup.

Systems are built to natively understand what things are.

HTML 5

HTML5 is
not presentational.

Rendering is done by
CSS.

This is what enables
adaptability
to various, and new,
contexts.

One HTML file
renders how you need it to
in print, online, on a phone,
on an e-reader, etc.

The **Open Web Platform** *is the basis for:*

All Modern Browsers.
*Used for online viewing of content
and the basis for many e-readers and platforms.*

Most Apps.
*Each app is programmed for a specific OS,
but they almost all use OWP technologies.*

Most E-Readers.
*Apple iBooks. Kobo. Nook. Kindle. Google Play.
VitalSource. Overdrive. Most proprietary
e-reading platforms (including journal hosts!).*

APEX
Content Solutions

NEWS FLASH:

The Web was not created for publishing.

NEWS FLASH:

The Web was not created for publishing.

NEWS FLASH:

The Web was not created for publishing.

It was mainly created for communication, finding stuff, and selling stuff.

W3C®

The W3C
has formed the

**Digital Publishing
Interest Group**

to explore how
Web technologies
can be improved
for publishing.

W3C Digital Publishing Interest Group

Main mission:

Document use cases
*for things publishers need to do
using Web technologies.*

Work with appropriate
W3C Working Groups
*to update the specifications
and standards.*

HTML

APEX
Content Solutions

Current Areas of Work in the DPIG:

Typography and Layout
Achieving print-quality rendering via CSS

Annotations and Social Reading
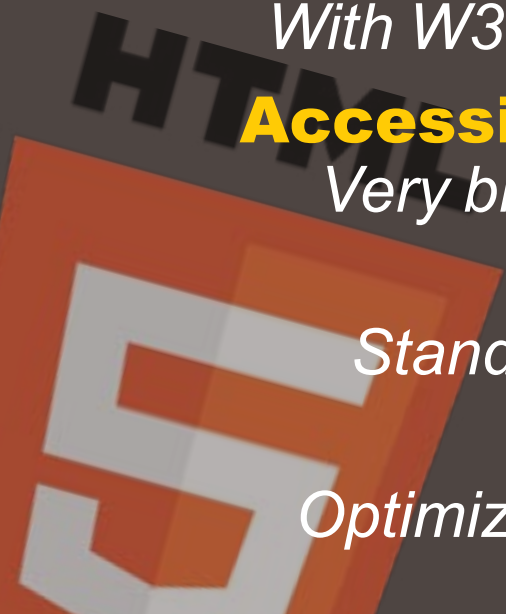*With W3C and EPUB Annotations WGs*

Accessibility and Personalization
Very broad, overlaps all other areas

Behavioral UI
Standardizing things like pop-ups

Metadata
Optimizing schema.org for publishing

Current Areas of Work in the DPIG:

Typography and Layout
Achieving print-quality rendering via CSS

Annotations and Social Reading
With W3C and EPUB Annotations WGs

W3C Open Annotations WG

*Building on the work of the W3C
Open Annotations Community Group,
on which the EPUB Annotations spec is based.*

Very complex issue, managing annotations and their provenance across time and documents.

DPIG Collaborating with PF WG of WAI

(Translation: Protocols and Formats Working Group of the W3C Web Accessibility Initiative)

First focus: Expand Structural Semantics for ARIA @role to add important publishing terms.
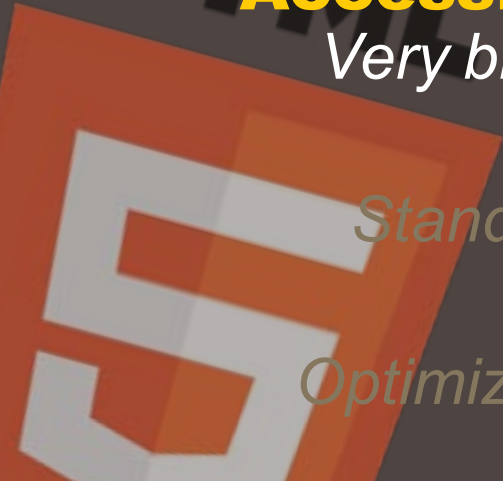
Accessibility and Personalization
Very broad, overlaps all other areas

Behavioral UI
Standardizing things like pop-ups

Metadata
Optimizing schema.org for publishing

Current Areas of Work in the DPIG:

Typography and Layout
Achieving print-quality rendering via CSS

Annotations and Social Reading
With W3C and EPUB Annotations WGs

Schema.org is "built in" to HTML5
Enables "semantic enrichment" (adding metadata) to granular components of HTML5 documents.

Metadata
Optimizing schema.org for publishing

STANDARDS THAT PEOPLE THINK ARE W3C STANDARDS BUT AREN'T, EXACTLY.

APEX
Content Solutions

Some Examples Important to STM

schema.org
Associates metadata with granular components, e.g. "address," "company name," "recipe ingredient."

Not a W3C standard; a collaboration between major search engines and browsers.

Enables "rich snippet"-like functionality, with "properties" that browsers natively understand.

Properties recently added for accessibility and education; BISG is working on ONIX in schema.org.

APEX
Content Solutions

Some Examples Important to STM

schema.org
Associates metadata with granular components

Schematron
Used for context-based QC, business rule conformance

XPath-based, very closely related to XSLT:
XSLT <u>transforms</u> *from one XML to another;*
Schematron <u>reports</u> *on an XML file.*

Key complement to parsing:
Schematron can flag errors or potential errors
that parsing can't detect.

Some Examples Important to STM

schema.org
Associates metadata with granular components

Schematron
Used for context-based QC, business rule conformance

ODRL
The Open Digital Rights Language:
designed to be both human- and machine-readable.

From a W3C "Community Group," not a W3C standard.

Describes "assets," "parties," "policies," "actions," "permissions," "prohibitions," "duties," "constraints," etc.

Some Examples Important to STM

schema.org
Associates metadata with granular components

Schematron
Used for context-based QC, business rule conformance

ODRL
The Open Digital Rights Language

JSON
*JavaScript Object Notation,
a lightweight data-interchange format;
simple subset of JavaScript, née EcmaScript.*

STANDARDS THAT ARE
BUILT ON WEB STANDARDS

APEX
Content Solutions

The Most Prominent Standard(s) in STM:

JATS/BITS (née NLM) XML
*is a pair of tightly coordinated tagsets—
JATS for journals and BITS for books.
JATS V 1.1d1 (NISO) ≈ BITS V 1.0 (NCBI):
below the chapter level, BITS markup = JATS markup.*

They incorporate other OWP standards:

MathML 3.0

Table Model based on
XHTML 1.1 Table Model

XInclude *for modularization*

APEX
Content Solutions

The Most Prominent Standard(s) in STM:

JATS/BITS (née NLM) XML
is a pair of tightly coordinated tagsets—
JATS for journals and BITS for books.

So the math and tables in JATS/BITS XML can already be HTML-conformant.

MathML 3.0

Table Model based on
XHTML 1.1 Table Model

XInclude *for modularization*

APEX
Content Solutions

EPUB 3 is basically packaged Web content.

It's for the interchange and distribution of digital publications and documents based on Web standards.

Content Documents are HTML5 or SVG.
As HTML5 evolves, EPUB 3 will keep pace with it.

HTML5 markup is expressed as XML.
*This provides more rigorous markup
and enables namespaces, e.g. @epub:type.*

Rendering is done via CSS.
You can have multiple stylesheets, plus responsive design.

Media and scripting align with HTML5.
*Any HTML <video> or <audio> works in EPUB 3,
and scripting uses JavaScript.*

EPUB 3 is built for accessibility
and aligns with Web accessibility standards.

What EPUB 3 adds to all that Web content:

A container and packaging spec.
EPUBs are designed to work online or offline.

A manifest and spine.
These document and organize all of the components.

Metadata.
Lots possible, little required; uses Dublin Core and <meta>.

Structural Semantics Vocabulary.
*Now you can specify that a <section> is a chapter,
and that an <aside> is a footnote.*

Lots more . . .
to make diverse components work as a publication.

And EPUB 3 continues to evolve.

EPUB 3.0.1 *(finalized June 2014)*
Integrates FXL, adds <collection>, expands semantics.

Indexes *(in final approval stage)*

Dictionaries & Glossaries *(in final approval stage)*

Multiple Renditions *(in final approval stage)*

Previews *(in final approval stage)*

Region-Based Navigation *(in final approval stage)*

Open Annotations *(draft spec, aligning with W3C WG)*

Go to* http://idpf.org/ongoing *to see the latest status.

Now we're getting to **a real ecosystem** based on **standards that are based on standards.**

So we can now build securely on this.

Now we're getting to a real ecosystem based on standards that are based on standards.

So we can now build securely on this.

A case in point:

THE EDUPUB PROFILE OF EPUB 3.

Specs for how to create an EPUB 3 that's optimized for education and interoperability in the educational ecosystem.

APEX
Content Solutions

The EDUPUB ALLIANCE
*is a collaborative effort between
the **IDPF** (EPUB for distribution and interchange),
IMS Global (QTI, LTI, Caliper Analytics, etc.),
the **W3C** (the Open Web Platform), and
BISG (the Book Industry Study Group).*

THE EDUPUB PROFILE OF EPUB 3.

Specs for how to create an EPUB 3 that's optimized for education and interoperability in the educational ecosystem.

APEX
Content Solutions

OWP: The Open Web Platform

HTML5
XML
CSS
SVG
MathML
ARIA
and
many
others

EPUB 3

Container
Package
Metadata
Content Docs
Resources
 (images, fonts,
 scripts, media)
Media Overlays
Structural Semantics
etc.

EDUPUB

A profile of EPUB 3 optimized for education.

It IS an EPUB.

Additions to the EPUB 3 Spec prompted by the development of EDUPUB:

The need for Interactive Widgets accelerated development of **EPUB Scripted Component** spec.

The need for Distributable Educational Objects prompted **EPUB Distributable Objects** spec.

IMS Global published Best Practices for using **Caliper Analytics**, **QTI**, and **LTI** with EPUB 3.

Need for educational vocabulary prompted expanded **EPUB Structural Semantics**.

Now, the ultimate in collaboration and convergence:

EPUB-WEB

Initially presented at Books in Browsers in October.

White Paper authored by Markus Gylling (IDPF CTO) and Ivan Herman & Ralph Swick (W3C)—an unofficial draft.

Articulates a long-range vision:
**An EPUB and a Website are
two different "states" of a single publication.**

*Identical experience in a browser or in an e-reader,
or on an app on a phone or tablet.*

Expected to take years, but is definitely do-able.

APEX
Content Solutions

IDENTIFIERS AND METADATA

are the oil that makes all this work smoothly.

IDENTIFIERS

ISBN

The International Standard Book Number

(Actually, Int'l Standard Product-Format-of-a-Book No.)

Update nearing WG approval, then to ISO.
No major changes.

Deprecates 10-digit ISBN: all ISBNs are 13-digit.

Aligns metadata better with ONIX.

New section on DOI and ISBN-A.

Clarifies assignment to digital and print products.

User Manual expected in December.

Spec is free & short; detail is in user manual.

APEX
Content Solutions

ISTC

The International Standard Text Code

The "work identifier" everybody needs but nobody seems to want.

Basic problem: supply chain sees "work" in Indecs terms; wider ecosystem thinks in FRBR terms.

Publishers track "their works" in their own systems.

Task force from RAs is looking at issues.

Scope and granularity of ISTC are not correct.

Stay tuned for revived activity.

APEX
Content Solutions

ISNI
The International Standard Name Identifier

**Identifies the "public identity"
of individuals and organizations.**

*Bowker, US RA, issues individual ISNIs (mainly authors);
about to release ISNIs in its ONIX feed.*

Ringgold issues organization ISNIs.

Must have sufficient corroboration to be issued.

Adoption is rapid: >8,000,000 so far.

2.33 million from Books in Print; .5 million are orgs.

Uptake strong in trade pub. and entertainment.

APEX
Content Solutions

ORCID
The Open Researcher and Contributor ID

Specifically for authors/contributors, mainly (but not exclusively) STM researchers.

Identity disambiguation like ISNI, but more about enabling an individual to maintain a public, accurate record of her contributions.

API services associated with metadata DB are key.

ISNI Interoperability tied to an EC proposal.

Rapidly gaining traction in STM; UI now available in 10 languages.

ISLI
The International Standard Link Identifier

Just getting started.

Registers a link between a source and a target and metadata about the link.

Agnostic as to what is linked and how it's executed.

Links can use ISBN, ISSN, DOI, ISRC, ISNI, etc.

Watch this space; not ready for prime time.

Spec is complete and RA proposals being evaluated.

Expect 2015 publication & RA implementation.

APEX
Content Solutions

METADATA

Thema
The International Standard for Subject Classification in the Global Book Trade

Collaborative initiative by many countries.

Thema provides a scheme across many languages.

National extensions enable Thema to "look local and act global."

1.1 released Dec. 1, 2014; updated annually.

Doesn't have to replace BISAC.

BISG recommends using both.

APEX
Content Solutions

JATS4R
JATS for Reuse

A brand-new initiative of the OA community.

Addressing problem of rich metadata used inconsistently.

The flexibility of JATS impedes interoperability.

Working to establish guidelines for how to tag metadata in journal articles to facilitate reuse.

Will conform to JATS.

LCC
The Linked Content Coalition

A new not-for-profit consortium of standards bodies and identifier and metadata registries: DOI Foundation, IPTC, NISO, EDItEUR, etc.

Building on work of 2012 LCC Project.

It's all about interoperability.

Encouraging use of highly interoperable identifiers, metadata, and messaging standards.

Streamline links between creators/rights holders and consumers and cultural institutions.

APEX
Content Solutions

Standards built on standards.

Collaboration.

Convergence.

It's working.

APEX
Content Solutions

Thanks!

Bill Kasdorf

bkasdorf@apexcovantage.com

+1 734 904 6252

@BillKasdorf