

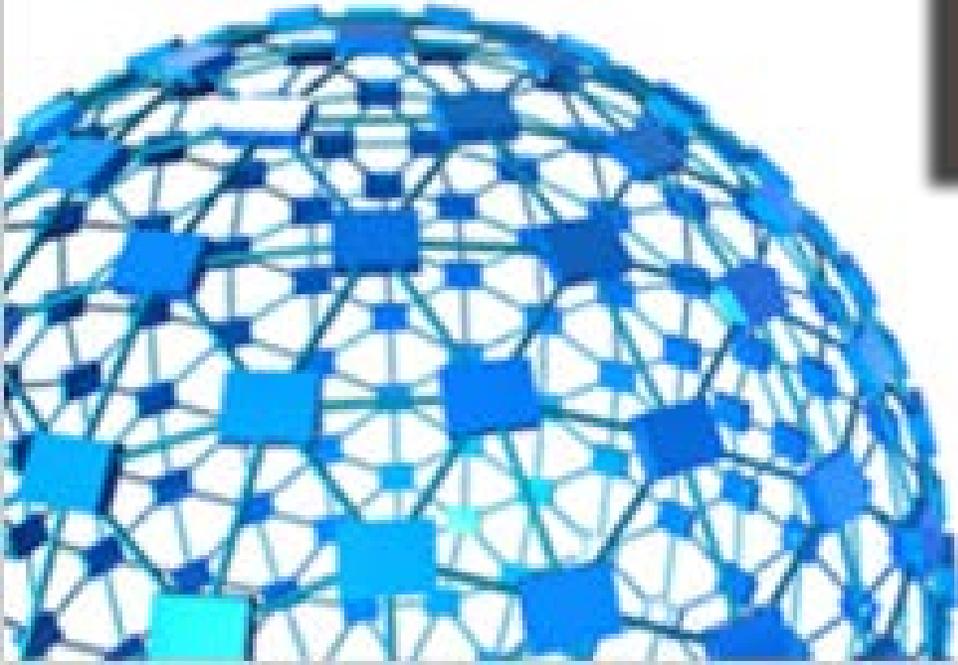


The Convergence Continues

An update on the evolution of our increasingly interoperable publishing ecosystem

Bill Kasdorf

*VP and Principal Consultant, Apex Content Solutions
Member of IDPF Board, EPUB 3 WG, W3C DPUB IG*



Not a news flash:

**We live in an
interconnected
world.**

INTERCONNECTEDNESS

leads to

INTERDEPENDENCE.

INTERCONNECTEDNESS

leads to

INTERDEPENDENCE.

This is not a bad thing:

*We don't have to start from scratch when
developing standards and building technologies.*

INTERCONNECTEDNESS

leads to

INTERDEPENDENCE.

This is not a bad thing:

*We don't have to start from scratch when
developing standards and building technologies.*

The new mandate:

INTEROPERABILITY.



INTERCONNECTEDNESS

leads to

INTERDEPENDENCE

which requires

INTEROPERABILITY

which leads to

CONVERGENCE

which is created by

COLLABORATION.

**The scholarly publishing ecosystem
is more interoperable
than that of any other sector.**

**The scholarly publishing ecosystem
is more interoperable
than that of any other sector.**

Some of the reasons why:

JATS/BITS XML

DOIs

CrossRef

Journal Hosting Platforms

Subscriptions + Open Access

The consumers are the creators

*This was achieved through
collaboration between publishers*
that are otherwise competitors . . .
. . . because **citation and access**
are essential to research.*

JATS/BITS XML
DOIs
CrossRef
Journal Hosting Platforms
Subscriptions + Open Access
The consumers are the creators

**Actually between vendors, libraries, aggregators, and lots of others too.*

This was achieved through collaboration between publishers that are otherwise competitors because citation and access are essential to research.

EIDR: Entertainment ID Registry
DOI-based identifier + metadata
CrossRef-like infrastructure + services

*Similar infrastructure developing in **entertainment** due to **collaboration** between the big players.*



**So we're
all set,
right?**

*We've
got it all
together. . . .*



Well, pretty much . . .

**. . . as long as we can
stay inside our
siloesd STM world.**



**The Grand Convergence
that is underway now
in the publishing ecosystem
is the result of
multidimensional collaboration.**

**The Grand Convergence
that is underway now
in the publishing ecosystem
is the result of
multidimensional collaboration.**

Cross-sector collaboration.

**The Grand Convergence
that is underway now
in the publishing ecosystem
is the result of
multidimensional collaboration.
Cross-sector collaboration.
Global collaboration.**

**The Grand Convergence
that is underway now
in the publishing ecosystem
is the result of
multidimensional collaboration.**

Cross-sector collaboration.

Global collaboration.

Standards based on standards.

**The Grand Convergence
that is underway now
in the publishing ecosystem
is the result of
multidimensional collaboration.**

Cross-sector collaboration.

Global collaboration.

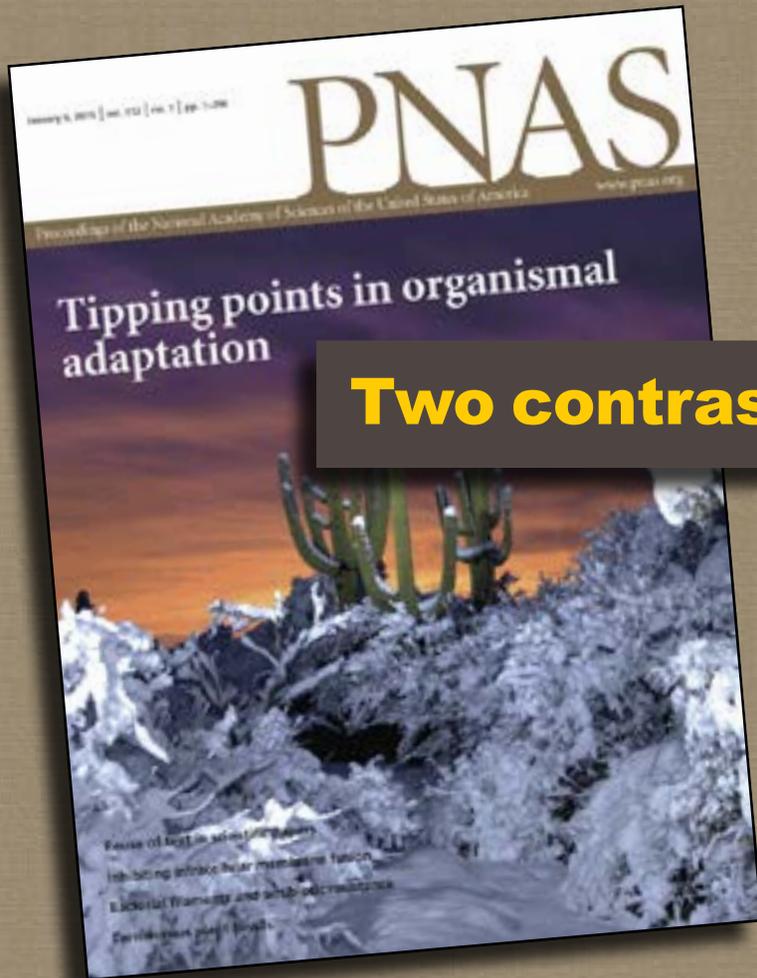
Standards based on standards.

Here are some examples.

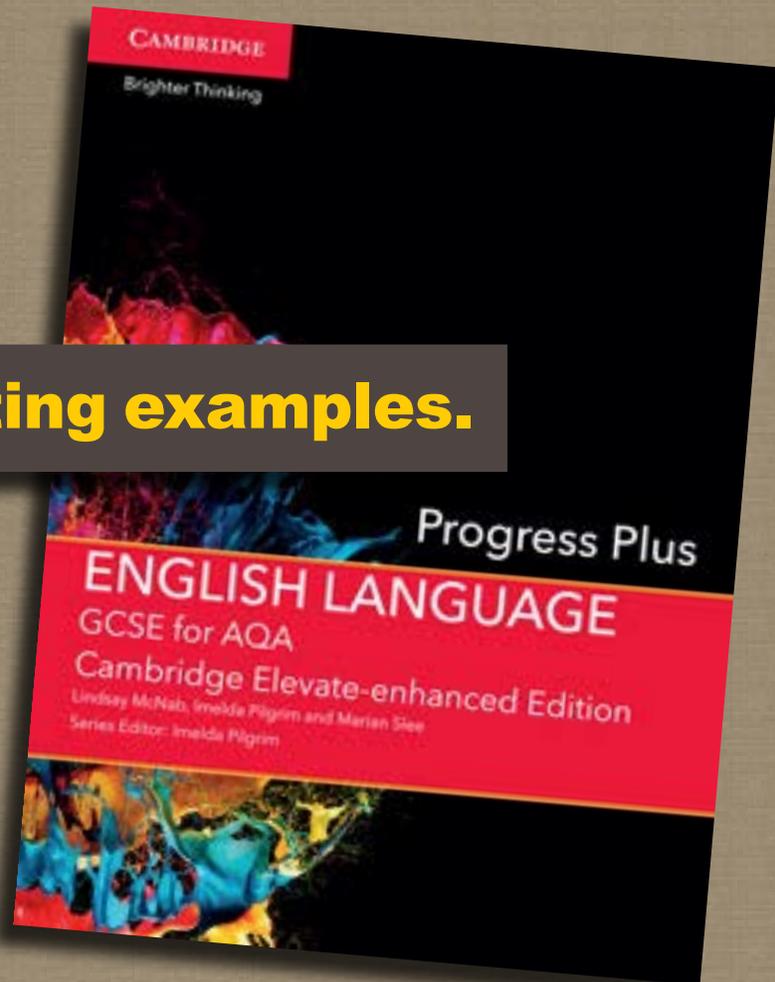


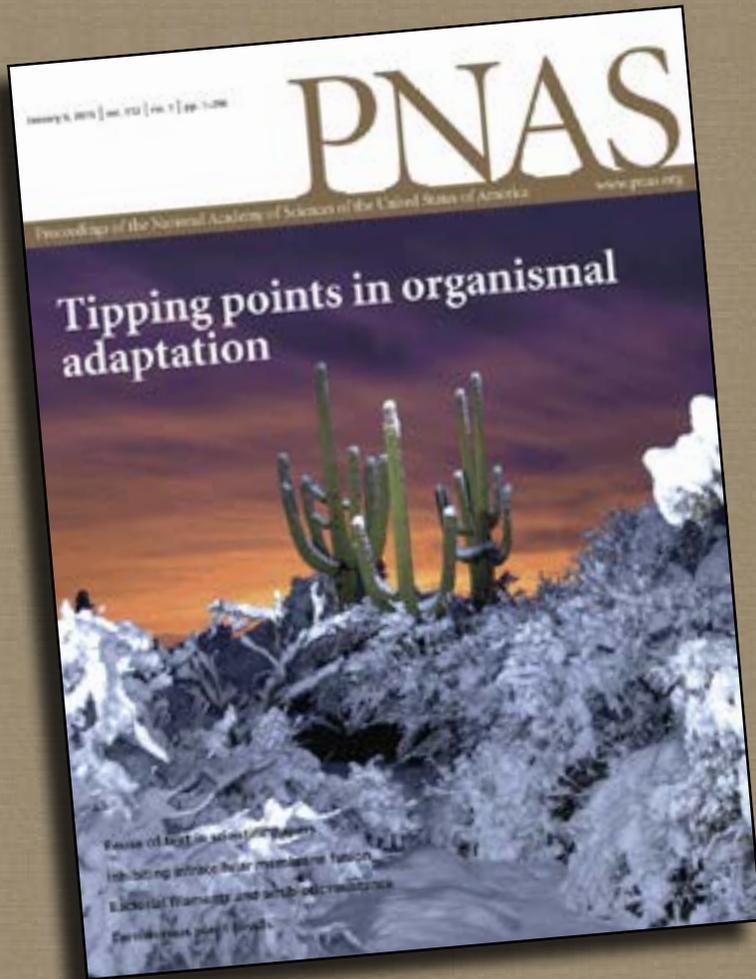
**Making rights expressions
interoperable.**

Rights



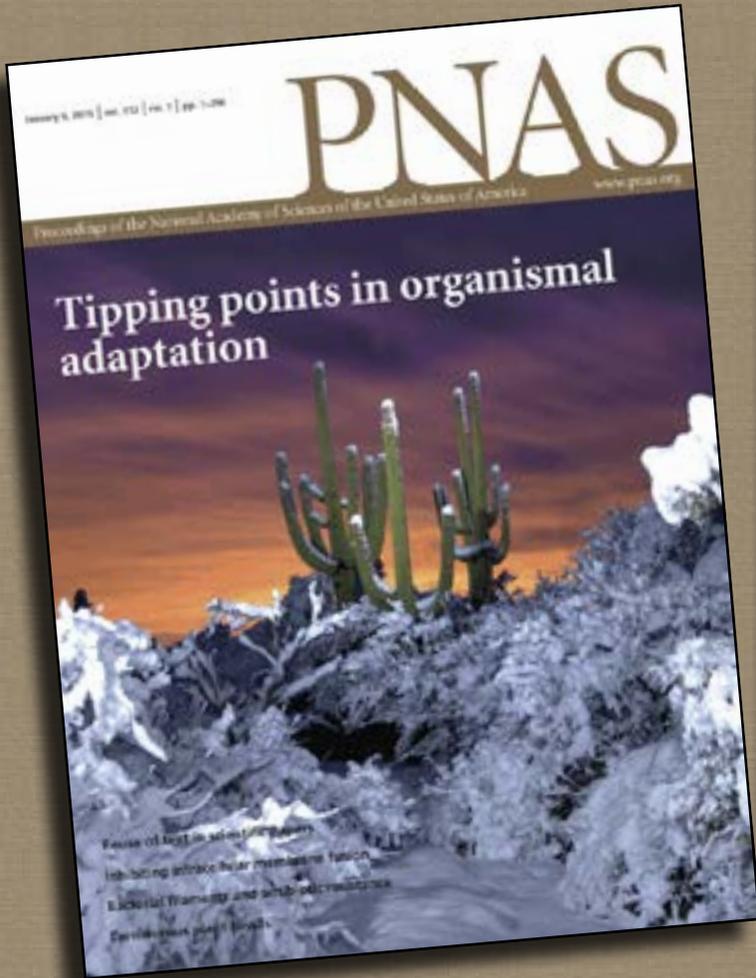
Two contrasting examples.





HYBRID JOURNAL

*Fundamental issue:
is this article open access
or not?*



A simple, elegant solution:

ALI, Access License and Indicators

Just two elements:

<free-to-read>

(with optional start and end dates)

<license-ref>

*(points to stable URI
for license terms, e.g.,
on platform or Creative Commons)*

ALI: NISO Access License and Indicators

NISO Recommended Practice RP-22-2015

*Incorporated in JATS XML w/ ali: prefix;
CrossRef already had LicenseRef,
added <free-to-read> in CrossRef API.*

**NISO WG was a collaboration
within the scholarly publishing sector:**

*CrossRef, PLoS, SPARC, EDItEUR, STM, ACS,
Indiana U. and U. of Birmingham Libraries, SSRN,
Wellcome, CCC, CC, Europeana, JISC, Elsevier*

ALI: NISO Access and License Indicators

Why was this so easy to do?

Very homogeneous ecosystem.

Virtually all journal access is online, article-based.

*Established support systems
(hosting platforms, CrossRef, academic libraries).*

*Only concerned with the article,
not components of the article.*

*Easy to incorporate in JATS metadata header,
and the metadata is in the asset.*

COMPLEX MULTI-SECTOR EXAMPLE:
A huge problem needing to be solved.

TEXTBOOK

(Print, ebook, and platform)

Could contain:

*Video from Condé Nast
Interview from the BBC*

Photo licensed from Getty

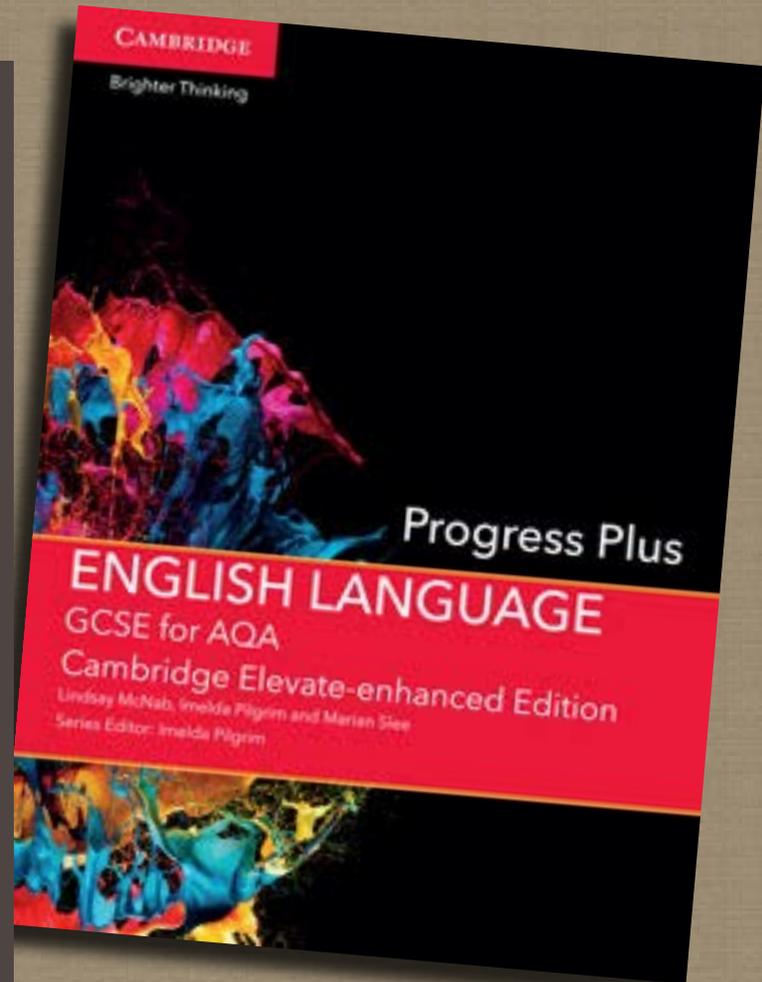
Article from PNAS

Commissioned illustration

Text from The Goldfinch

Educator-created content

Student-created content



TEXTBOOK

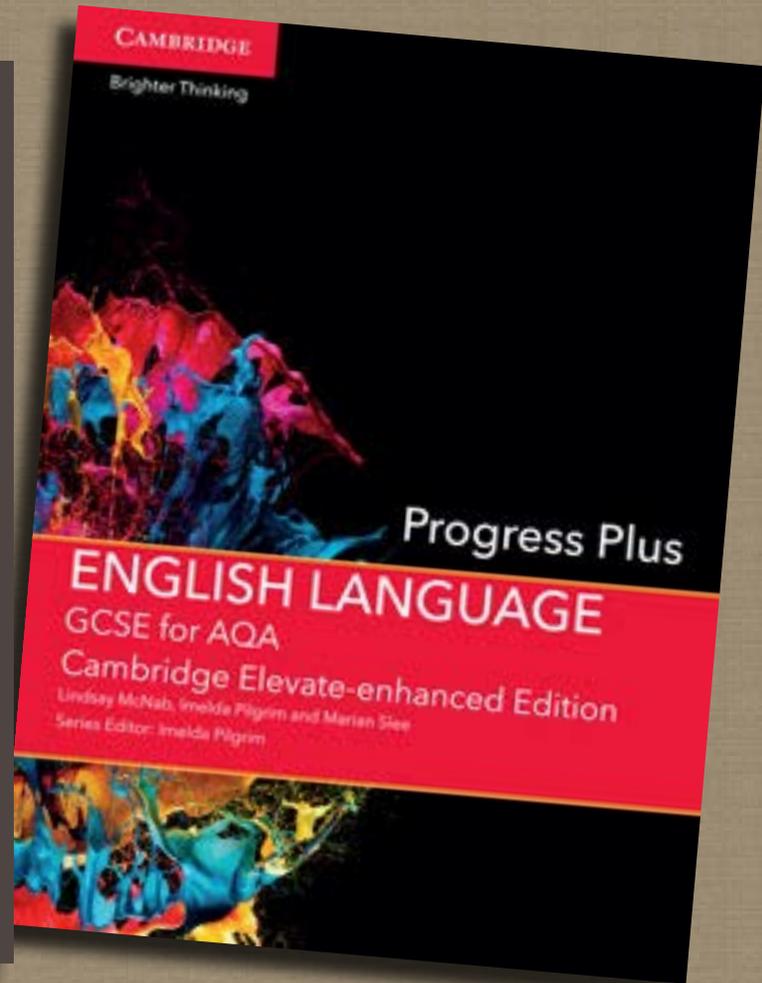
(Print, ebook, and platform)

What *rights* did they obtain for each of those assets?

What *formats* are the assets provided as?

What *metadata* came with them?

Do any of them have *embedded assets*, with *embedded rights*?



Not simple!

**Those assets are all from
different sectors.**

**The file formats may all be different.
They may not just plug right in.**

**The sectors don't all use the same
rights metadata, and most of it
is not machine-resolvable.**

The metadata isn't usually in the asset.

**And a big textbook/platform has
hundreds of such assets.**

OMG! OMG! OMG!

What a mess!



**This calls for
MULTI-SECTOR COLLABORATION.**



BISG

EDITEUR

W3C

IPTC

PLUS

Let's look at rights activities in some standards groups.

BISG

**Focus on
authors
and agents.**

**Royalty
tracking**

*is a
nightmare,
especially
for ebooks,
especially
for agents.*



BISG

**Focus on
authors
and agents.**

**Controlled
vocabulary:**

*Rich, modular,
structured.*

*Designed for
machine/human
interoperability.*



BISG

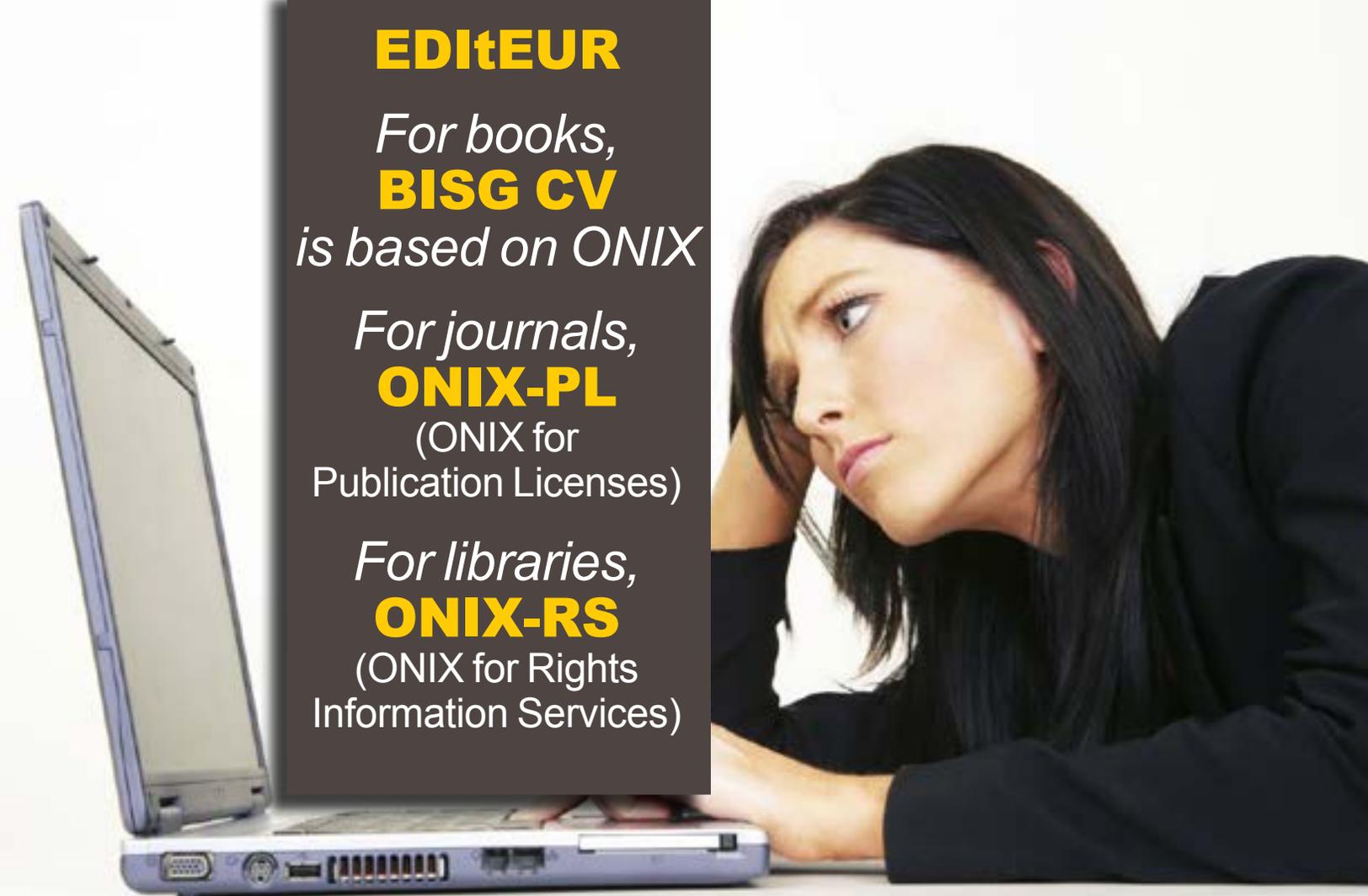
**Focus on
authors
and agents.**

**Payment
spec:**

*Very complex,
detailed model
aligned with CV.*

*Based on CCC
payment spec.*





EDITEUR

For books,
BISG CV
is based on ONIX

For journals,
ONIX-PL
(ONIX for
Publication Licenses)

For libraries,
ONIX-RS
(ONIX for Rights
Information Services)

**Books and journals
can live for now with having to
look up and read the licensing terms.**

**Books and journals
can live for now with having to
look up and read the licensing terms.**

The news industry and magazines can't.

**Books and journals
can live for now with having to
look up and read the licensing terms.**

The news industry and magazines can't.

An example of why:

The Associated Press produces

250,000 assets EVERY DAY.

**And they have to manage the rights
associated with all of them.**

ODRL

(Open Digital Rights Language)

*Framework for the
machine-readable
transmission of
rights metadata.*

Enables expressing:

*Policy Types
Actions
Constraints
Parties and Roles
Relations*





**Wait . . . is that a real
Web standard?**



**Wait . . . is that a real
Web standard?**

**Well, not quite . . .
but stay tuned.**

*It's from a CG, not a WG,
but it's about to get on a "rec
track" to become a formal
W3C Recommendation:*

**Open Permissions
and Obligations
Expression**



IPTC RightsML

Profile of ODRL.

*Machine processable,
embeddable in assets
as XML, RDF, or JSON.*

*And the IPTC
is developing
**open-source
RightsML software.***



Photo Industry: PLUS Coalition

(Picture Licensing Universal System)

*Very broad membership
(creators, distributors,
users, RROs, etc.) in
154 countries.*

*Collaborating with IPTC,
LCC, ODRL, ISNI,
EIDR, Idealliance, etc.*

*+ IPTC Photo Metadata
is the most widely used.*

What do all these have in common?

The Open Web Platform.

**Web technologies and
web services are
the context of the convergence.**

**They're what enable
INTEROPERABILITY.**

The Open Web Platform

*Over 100 W3C specs:
XML, HTML, CSS, SVG,
MathML, many more.*

HTML5

The structure.

CSS3

The styling.

JavaScript

The functionality.



**Remember:
it's not just about your website.**

**The Open Web Platform
is fundamental
to the how content is**

Developed

Discovered

Delivered

HTML





EPUB 3 is basically packaged Web content.

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

It's the free, open, accessible, non-proprietary standard for delivery of publications based on the OWP.

Updating EPUB.

EPUB 3.1

EPUB 3.1

Backwards-compatible update

New spec format, better integrated

Deprecating unused features of EPUB 3.0

The serialization issue: HTML vs. XML, @role

New approach to metadata

Possible new Core Media Types

Improved and stricter accessibility support

First draft in January, final by Oct. 2016

EPUB 3.1

Important new concept under consideration:

Browser-Friendly Format

*Specification for how to create an EPUB
with no features incompatible with browsers.*

Will not include any deprecated features.

Step toward more complete alignment with OWP.

Surveys re: metadata and deprecations

[Deprecation survey closes in two weeks, Dec. 15!]

**Progress on the
EPUB+Web convergence.**

PWP

The Portable Web Publication Vision:

**ONE FILE FOR BOTH
ONLINE AND OFFLINE USE.**

The same content in two different “states”:

Packaged for offline and archival use;

Unpackaged for online or cached use.

PWP: Portable Web Publications

The Portable Web Publications vision:

“Reduce the current format- and workflow-level separation between offline/portable and online (Web) document publishing to zero.”

These are merely two dynamic manifestations of the same publication. Links, annotations, cross-references, etc. have to “just work.”

Publishers and users can employ either of these publishing and consumption modes.

PWP: Portable Web Publications

Latest public working draft published Nov. 26, 2015

Explicit focus on “Scholarly Journal & STM Publishers”

Identifiers are a huge issue

How to associate multiple documents with one URL?

*Need two identifiers: a generic, stable URN (name)
(which can use an ISBN, DOI, etc.) and
a URL (locator) that is essentially an HTTP URI.*

Then need to identify components and fragments.

**And while we're at it,
let's make math render better.**

MathML

MathJax

MathJax 2.6

*In public beta, final release planned this month
Improved accessibility, esp. for MathML-enabled AT*

*New “CommonHTML” format:
much faster, stable across browsers*

Significant Changes Coming in 2016

*Based on CommonHTML; improved modularity
Can't assume browsers will implement MathML well
See MathJax white paper coming out this month*

MathJax

MathJax Semantic Enrichment Project

Initial funding from Alfred P. Sloan Foundation

Enriches Presentation MathML with semantic info

Two demonstration applications:

Responsive equations (e.g., for mobile)

Math Accessibility Solution (on-the-fly speechevent)

Call for Sponsors

Sponsors are involved in development

Funding needed for accessibility and semantics work

**Making publications
“Born Accessible”**

Accessibility

Toward a “Baseline for Accessibility”

Imminent “refresh” of Section 508

Expected to align with WCAG 2.0 levels A and AA (Web Content Accessibility Guidelines): international; technology-neutral; apply to web, software, documents.

Updating WAI-ARIA 1.1 for publishing

Adding @role attributes need for publications

“EPUB Means Accessibility”

EPUB 3.1 requirements will be more explicit and strict.

BISG Quick Start Guide to Accessibility

Clear overview plus in-depth info; publishing this month.

**Sharing information
about research.**

SHARE

SHARE: SHared Access Research Ecosystem

“SHARE is building a free, open, data set about research and scholarly activities across their life cycle.”

*A new higher education initiative—
universities (AAU + APLU), libraries (ARL), plus
CrossRef, NISO, CHORUS, Portico, ORCID, etc.*

*Huge repository of metadata about research:
data management plans, grant proposals, preprints,
journal articles, data repository deposits.*

First project: SHARE Notify (public beta April 2015).

**Creating an interoperable
annotation ecosystem.**

Annotation

“Annotating All Knowledge”

Coalition being formed by Hypothes.is

Just announced today

Coalition of over 40 scholarly publishers, platforms, and technology partners: JSTOR, PLOS, HathiTrust, CrossRef, HighWire, Wiley, & many more

Goal: Establish a common framework for scholarly collaboration based on open standards

Building on W3C Open Annotations WG work; expect several years work, all open & standards-based

**Amping up the responsiveness
of the Web.**

AMP

AMP: Accelerated Mobile Pages

Open-source OWP-based Google project to make the web faster and more responsive.

Working with most big news organizations (e.g., NYT), magazine pubs, and CMS platforms (e.g., WordPress).

Publisher or partner creates and hosts AMP-HTML.

Google (or other platform) caches and serves.

Speeds delivery to laptops, tablets, and phones.

JavaScript-based; content loads almost instantly.

Framework is available now; see www.ampproject.org.

Expect formal launch Q1 2016.

And even some progress on ISTC!

ISTC

ISTC: International Standard Text Code

Attempt to provide a “work identifier.”

Has existed since 2009 (ISO 21047).

Little or no takeup; currently being completely rethought.

Problem: What do you mean by “work”?

Solution: Enable flexible “clustering.”

Enables grouping or distinguishing based on geography, language, abridgement, editions, etc.

The Registration Agency would manage the clustering and associated metadata and issue the ISTCs.

Resources

RIGHTS

NISO Access and License Indicators: <http://www.niso.org/workrooms/ali/>

BISG Rights Controlled Vocabulary: <https://www.bisg.org/rights>

ONIX-PL: <http://www.editeur.org/21/ONIX-PL/>

ONIX-RS: <http://www.editeur.org/124/ONIX-RS/>

ODRL 2.1: <http://www.w3.org/ns/odrl/2/ODRL21>

Open Permissions and Obligations Expressions: <http://w3c.github.io/ole/charter.html>

IPTC RightsML: <https://iptc.org/standards/rightsml/>

PLUS Coalition: <https://www.useplus.com/aboutplus/system.asp>

EPUB and PWP

EPUB 3.1 Workplan: <http://www.idpf.org/workplans/2015/epub/>

Portable Web Publications: <http://www.w3.org/TR/2015/WD-pwp-20151126/>

MathJax

MathJax 2.6: <http://docs.mathjax.org/en/v2.6-latest/whats-new-2.6.html>

Semantic Enrichment: <https://github.com/mathjax/MathJax/wiki/Semantic-Enrichment-project>

Resources

ACCESSIBILITY

BISG Quick Start Guide to Accessibility:

<https://www.bisg.org/publications/quick-start-guide-to-accessibility>

Section 508 Refresh: <http://www.access-board.gov/guidelines-and-standards/communications-and-it/about-the-ict-refresh>

SHARE

<http://www.share-research.org/>

AMP

Accelerated Mobile Pages: <https://www.ampproject.org/>

ISTC

[URL to come; current draft not yet public]

Annotating All Knowledge

Dec. 1 2015 Press advisory, with many useful links: https://docs.google.com/document/d/1_Xy-OPnn20SK2oqXe6VSJ8BHBx7m6r7c2Qc_Be9Y990/edit

Thanks!

Bill Kasdorf

bkasdorf@apexcovantage.com

+1 734 904 6252

@BillKasdorf

