



# EPUB 3.3, a W3C standard with a test suite

Ivan Herman

Digital Publishing Technical Lead

W3C

[ivan@w3.org](mailto:ivan@w3.org)

## Digital Publishing Summit

**June 3, 2022, Madrid, Spain**  
Casa del Lector - Paseo de la Chopera, 14

Where technical innovation  
meets publishing needs



# A bit of history...

---



- 2017: EPUB 3.1, last version defined by IDPF
  - had backward compatibility issues
- 2017: merger of W3C and IDPF, IDPF becomes a separate “activity” at W3C under the banner of “Publishing@W3C”
- 2018: EPUB 3.2, published as a W3C Community Group Note
  - started by EPUB 3.1, but rolling back the compatibility issues: ***any valid EPUB 3.01 document is a valid EPUB 3.2 document.***
- 2021: creation of the W3C EPUB 3 Working Group, with the goal of publishing a W3C Recommendation (informally, a “Web Standard”) for EPUB 3.3:
  - going through a rigorous W3C process
  - thorough editorial review
  - very few new features

# “Rigorous W3C process”

---



- “Horizontal Reviews”
  - experts in accessibility, internationalization, security, or privacy review the documents and raise issues if necessary
  - working groups are *required* to take those in consideration
- Testing
  - the process includes a testing phase, a.k.a., “Candidate Recommendation”
  - there *must* be a public test suite and there *must* be (at least) two independent implementations for every normative feature
- Editorial requirements (format, terminology, etc.)



**The best news of all**



**Every valid EPUB 3.2 document  
is also valid EPUB 3.3!**



**Every valid EPUB 3.2 document  
is also valid EPUB 3.3!**

**...and EPUBCheck will be released with full support for  
EPUB 3.3**



# New technical features (only a few...)

## Digital Publishing Summit

**June 3, 2022, Madrid, Spain**  
Casa del Lector - Paseo de la Chopera, 14

Where technical innovation  
meets publishing needs



# It was not the goal to add many new features...

---

- New core media types:
  - WebP (modern image format for the Web)
  - OPUS (open source audio codec)
- HTML's `dir` attribute to display bidirectional texts in the package document
  - e.g., to display the following title text correctly: “!CSS: הרפתקה חדשה!”
- New accessibility recommendations for page lists and page numbering
- Add `refines` attribute for accessibility metadata fields:

```
<meta property="a11y:certifiedBy" id="certifier">Accessibility Testers Group</meta>  
<link rel="a11y:certifierReport" refines="#certifier"  
      href="https://example.com/a11y-report/">
```

# It was not the goal to add many new features...

---

- The accessibility specification is now conform to the European Accessibility Act
  - a separate note has been published to map the two documents
- Some technical additions to ensure security
  - incorporate the “(Web) origin” concept for EPUB
    - i.e., adopt the core notion of the Web security model for scripts within EPUB content documents
  - properly specify the “root” URL for a package
    - to avoid security risks if an author uses relative URLs of the sort:  
`../../../../leaking/into/filesystem`



# Editorial changes

## Digital Publishing Summit

**June 3, 2022, Madrid, Spain**  
Casa del Lector - Paseo de la Chopera, 14

Where technical innovation  
meets publishing needs



# Major reorganization of the documents

---

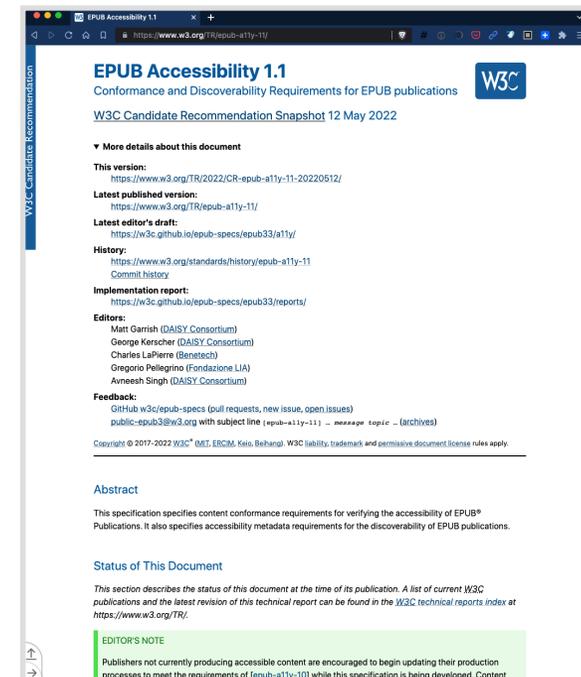
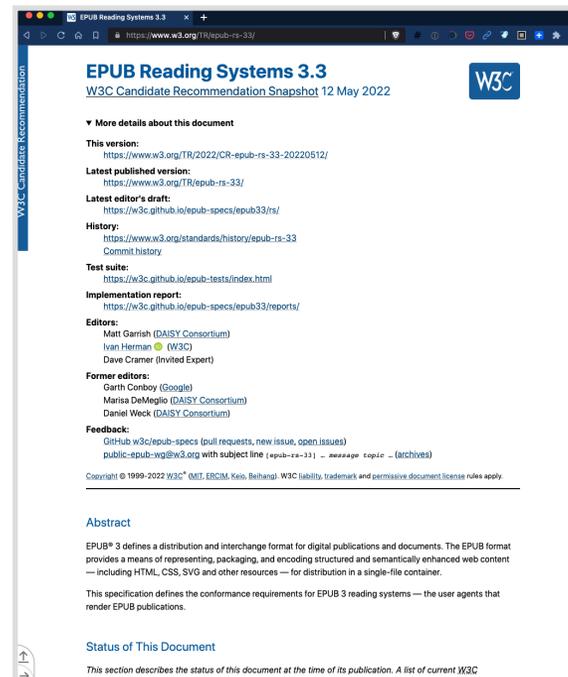
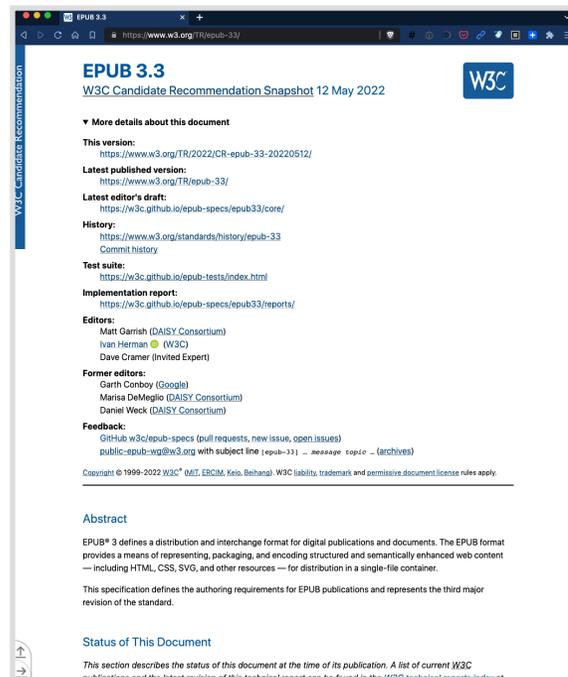


- This was the document suite for 3.2:
  - EPUB 3.2
  - EPUB Packages 3.2
  - EPUB Content Documents 3.2
  - EPUB Open Container Format 3.2
  - EPUB Media Overlays 3.2
  - EPUB Accessibility 1.0

# Major reorganization of the documents



- The three parts:
  - “Core” for content creators
  - “Reading Systems” for user agents and their implementors
  - “Accessibility 1.1” is now integral part of EPUB 3.3 (just like Web standards like WCAG 2.0)



# New privacy and security sections

---



- Added informative privacy and security sections for the the Core and the Reading Systems documents
  - help content authors and Reading System implementers to avoid privacy or security issues
  - include “threat models” in the documents, i.e., what are the threats in EPUB; e.g.,
    - scripting
    - compromised or malicious remote resources
    - phishing/spoofing
    - collection of user data
    - user-generated content

many of these are identical to the threats in Web pages, others are EPUB specific



Testing, testing!

## Digital Publishing Summit

June 3, 2022, Madrid, Spain  
Casa del Lector - Paseo de la Chopera, 14

Where technical innovation  
meets publishing needs



# EPUB interoperability

---



- Interoperability has long been a state we desired but had no real information on
  - lot of anecdotes (“this book works fine on Apple Books, but not on Kobo...”)
  - very little published, referenceable material
  - IDPF had a test suite, but incomplete and not actively maintained any more
- The new test suite for EPUB will provide some of that information
  - we can use to help inform future publishing standards

- A large, developing set of test
  - each test is a small EPUB 3.3 document, concentrating on *one* normative feature
  - each document describes, as part of its content, what it means to “pass” the test
  - reading systems are supposed to run the test and provide a file with yes/no answers on whether it passes or not
- There are, at the moment, around 130 tests and growing
- The goal is (per W3C process) that each test must be passed by at least two, *mutually independent* implementations
  - if that is not the case, the feature cannot be validated as normative

# There is a list of available tests



Id	Req	Title	Description	Specs	Ref
<a href="#">cnt-css-support</a>	must	Acid2 test	The Acid2 test, ported from <a href="http://acid2.acidtests.org">acid2.acidtests.org</a> , demonstrates broad CSS support.	(1)	<input type="checkbox"/>
<a href="#">cnt-mathml-support</a>	must	Support for Presentation MathML	A Presentation MathML equation is displayed as part of the XHTML content.	(1) (2)	<input type="checkbox"/>
<a href="#">cnt-svg-css-inclusion</a>	must	CSS applied to included SVG	CSS changes the color of an SVG image that is embedded in the XHTML content by inclusion.	(1)	<input type="checkbox"/>
<a href="#">cnt-svg-css-reference</a>	must	CSS not applied to referenced SVG	CSS does not change the color of an SVG image that is embedded in the XHTML content by reference.	(1)	<input type="checkbox"/>
<a href="#">cnt-svg-embedded</a>	must	SVG embedded in XHTML content	An SVG image is embedded in the XHTML content (as inline SVG).	(1)	<input type="checkbox"/>
<a href="#">cnt-svg-support</a>	must	SVG content document	An SVG image is displayed as a content document (not as part of an XHTML file).	(1)	<input type="checkbox"/>

# Tests and specification are mutually linked

Id	Req	Title	Description	Specs	Ref
<a href="#">cnt-css-support</a>	must	Acid2 test	The Acid2 test, ported from <a href="#">acid2.acidtests.org</a> , demonstrates broad CSS support.	(1)	<input type="checkbox"/>
<a href="#">cnt-mathml-support</a>	must	Support for Presentation MathML	A Presentation MathML equation is displayed as part of the XHTML content.	(1) (2)	<input type="checkbox"/>
<a href="#">cnt-svg-css-inclusion</a>	must	CSS applied to included SVG	CSS changes the color of an SVG image that is embedded in the XHTML content by inclusion.	(1)	<input type="checkbox"/>
<a href="#">cnt-svg-css-reference</a>	must	CSS not applied to referenced SVG	CSS does not change the color of an SVG image that is embedded in the XHTML content by reference.	(1)	<input type="checkbox"/>
<a href="#">cnt-svg-embedded</a>	must	SVG embedded in XHTML content	An SVG image is embedded in the XHTML content (as inline SVG).	(1)	<input type="checkbox"/>
<a href="#">cnt-svg-support</a>	must	SVG content document	An SVG image is displayed as a content document (not as part of an XHTML file).	(1)	<input type="checkbox"/>

## § 5.1.2.1 MathML

To support MathML [[mathml3](#)] embedded in [XHTML content documents](#), a reading system:

- **MUST** be an [input-compliant processor](#) for [Presentation MathML](#), as defined in the [[mathml3](#)] specification.
- **MAY** support visual rendering of Presentation MathML.
  - ▾ tests: 1
    - [cnt-mathml-support](#)
- **MAY** support rendering of [Content MathML](#) found in [annotation-xml elements](#) [[mathml3](#)].

### NOTE

Reading systems may choose to use third-party libraries such as [MathJax](#) to provide MathML rendering.

The goal is to have a full coverage of all normative statements by this summer

# MathML test example

---



- Per specification, a reading system must implement MathML
- There is a test (“cnt-mathml-support”) for this. Its (only) content document looks roughly as follows:

```
<html xmlns="http://www.w3.org/1999/xhtml" ...>
  ...
<body>
  <p>
    Test passes if a math equation appears below,
    including superscript (not plain text).
  </p>
  <math xmlns="http://www.w3.org/1998/Math/MathML">
    <mrow>
      <mi>x</mi>
      <mo>=</mo>
    </mrow>
    ...
  </math>
</body>
```

# MathML test example

Test passes if a math equation appears below, including superscript (not plain text).

$$x = (a + b)^2$$

Support for Presentation MathML [EPUB] — E-book viewer

Test passes if a math equation appears below, including superscript (not plain text).

$$x = (a + b)^2$$

Test passes if a math equation appears below, including superscript (not plain text).

$$x = a + b^2$$

Test passes if a math equation appears below, including superscript (not plain text).

$$x = (a + b)^2$$

# Language setting example

---



- Per specification the language set in the package document (e.g., “fr”) must not be automatically applied on the content document
- This is what test “pkg-lang\_but\_non\_content” is all about
- It has this setting in the package document:

```
<?xml version="1.0" encoding="UTF-8"?>
<package ... xml:lang="fr" ...>
  <dc:language>fr</dc:language>
  ...
</package>
```

but the content document does not have a language setting (i.e., it is considered to be “unknown”)

# Language setting example



Test passes if the following quotation is surrounded by the pair of "" and "" quotation marks (used in English) as opposed to the pair of «' and '» characters (used in French).

«Le mieux est l'ennemi du bien»

Test passes if the following quotation is surrounded by the pair of "" and "" quotation marks (used in English) as opposed to the pair of «' and '» characters (used in French).

"Le mieux est l'ennemi du bien"

Test passes if the following quotation is surrounded by the pair of "" and "" quotation marks (used in English) as opposed to the pair of «' and '» characters (used in French).

«Le mieux est l'ennemi du bien»

Test passes if the following quotation is surrounded by the pair of "" and "" quotation marks (used in English) as opposed to the pair of «' and '» characters (used in French).

"Le mieux est l'ennemi du bien"

# The test system generates a report



- This is based on what the implementations report about themselves

Id	Req	Android (PWA)	iOS (PWA)	Web (Chrome)	Web (Safari)
cnt-css-support	must	pass	pass	pass	pass
cnt-mathml-support	must	fail	pass	fail	pass
cnt-svg-css-inclusion	must	pass	pass	pass	pass
cnt-svg-css-reference	must	fail	fail	fail	fail
cnt-svg-embedded	must	pass	pass	pass	pass
cnt-svg-support	must	pass	pass	pass	pass

# What the test suite is *not*

---

- *The test suite concentrates on EPUB-specific features only*
- HTML, CSS, or SVG features are not covered
  - the recommendation relies on HTML, CSS, or SVG
  - these standards have their own (huge!) test suites, constantly growing
  - most of the new reading systems rely on some form of a Web “core” (a.k.a. WebView) that can be trusted on implementing those specifications properly



What about  
unimplemented features?

## Digital Publishing Summit

June 3, 2022, Madrid, Spain  
Casa del Lector - Paseo de la Chopera, 14

Where technical innovation  
meets publishing needs



# If you build it... they might not come

---



- Early versions of EPUB3 were optimistic, included features that sounded necessary and important to publications, but were not present in reading systems or tools. “If we build it, they will come” was the philosophy.
- More than 10 years later, many have been implemented, but some have not 😞
- Testing will allow to see where things stand but experiences are already clear for some features
- The work intend to bring this problems to the open to help authors to produce interoperable content

# Removing unused features

---



- Three types of unused features have been identified:
  - under-implemented features
    - there may be one or more implementations, but are not common and should be considered risky for content creators
    - marked as such if there is a single (or no) implementation that reports having implemented it
  - deprecated features
    - the working group no longer recommends to use them, as they have limited to no support
    - epubcheck warns about its usage
    - example: the `epub:switch` element
  - legacy features
    - retained for backwards compatibility with previous major versions of EPUB, but may not be supported by reading systems
    - example: NCX format for table of content

# Examples for (potentially) under-implemented features

---

- `rendition:flow`
  - the ability to tell a reading system to display content as vertically scrolled.
  - scrolling feature is implemented in many reading systems, but it is usually at the user discretion rather than creators'
- Manifest fallbacks
  - allow content creators to provide an alternate version of a file if the main one does not render
  - so far it appears that this feature is under-utilized, and also under-implemented in reading systems

# Removing features from the core specifications

---

- Some features have already been taken out of the standards due to next to no usage and implementations. E.g.:
  - multiple renditions
  - text-to-speech (i.e., SSML, pronunciation lexicons)
  - structural semantics vocabulary

These are published as separate, informal notes for now

- It is o.k. to use these features, but authors have to be aware of a limited availability in reading systems



# Some closing words

## Digital Publishing Summit

June 3, 2022, Madrid, Spain  
Casa del Lector - Paseo de la Chopera, 14

Where technical innovation  
meets publishing needs



# EPUB 3.3 is here

---



- The WG considers that technical changes are done
  - you can look at the document for content creation, it may be more helpful than the old one
  - you can use the latest version of epubcheck, it will help for EPUB 3.3
- The main emphasis from now on is testing
- Goal is to get a complete test suite by the summer and have suitable testing results by the end of the year
- The final standard should be published end of Winter 2023

# How can you help?

---



- Review and comment on the specifications
- Test the tests suite
  - each test is a separate EPUB 3.3 publication; it is important to ensure they are correct...
  - *provide new tests*, e.g., if a feature is not fully tested yet
- If you are a reading system implementer, run the test suite and publish a report
  - the more the merrier...

# Documents to look at

---

- Look at the WG Publication Status page for a list and references of all documents:
  - <https://www.w3.org/publishing/groups/epub-wg/PublStatus>
- Main documents on tests:
  - Test descriptions: <https://w3c.github.io/epub-tests/>
  - Implementation reports: <https://w3c.github.io/epub-tests/results>
  - How to contribute new tests: <https://w3c.github.io/epub-tests/contributing>



These slides can be downloaded from:  
<https://www.w3.org/2022/Talks/PublishingSummit-IH/>



These slides can be downloaded from:  
<https://www.w3.org/2022/Talks/PublishingSummit-IH/>

Thank you for your attention!