



EPUB



EPUB/A

Neil Jefferies, OPF



Preserving eBooks Project

Multiple partners across publishers, libraries and repositories

- Graham Bell (EDITEUR)
- David Cirella & Euan Cochrane (Yale)
- Gareth Cole (COPIIM)
- Rebecca Cook (Wiley)
- Sarah Fricker (IOPP)
- Matt Garrish (DAISY)
- Thib Guicherd-Callin (LOCKSS)
- Clare Hodder (rights2)
- Claire Holloway (OCLC)
- Neil Jefferies (OPF & Oxford)
- Ruth Jones
- Roxanne Missingham (Australian National University)
- Lynette Owen
- Sylvia Pegg & Becky Roberts (RSC)
- James Phillpotts (OUP)*
- Wendy Reid (Rakutan)
- Denis Shannon (U of Wyoming)
- Tzivya Siegman
- Alicia Wise (CLOCKSS)

3 Key Outcomes

1. Advocacy material aimed at publishers, [“Preserving Books for Future Generations”](#)
2. Prototype for *Keepers’ Registry* for eBooks
 - a. Add preservation details to ONIX (code 47/48) and MARC (field 857) feeds from publishers to WorldCat, GOBI, Nielsen Book Data, etc.
 - b. Also, KBART feeds from preservation repositories such as CLOCKSS and Portico
3. ePUB/A Specification



EPUB vs PDF

PDF Reading Assumptions

- Print-oriented - based on a cut-down PostScript PDL processor
- Focus on consistent appearance across devices/formats
- Text extractability and re-usability are not considerations

EPUB Reading Assumptions

- Device-oriented - designed for a variety of e-readers - PDA's, phones, tablets
- Focus on good reading experience regardless of device/screen size
- Reflowable text is easy to extract

Feature creep

- Tendency to acquire browser-like features such as JavaScript



EPUB/A Process

What does /A (“for Archiving”) mean?

- Preservation-friendly, easy to migrate
- Inspiration from PDF/A

Strict subset of EPUB

- Additional requirements would be a barrier to adoption
- Avoid external dependencies (fonts and content, links are OK)
- Avoid embedded executable content (hard to migrate, security risks)
- Avoid bitmapped and physical text layouts (compromises text extraction)

EPUB is a multiple standard

- W3C EPUB 3.3 Last updated Oct 2025
- ISO/IEC 23736-1:2020 EPUB 3.01
- ISO/IEC TS 22424-1:2020 EPUB Preservation principles
- ISO/IEC 23761:2021 EPUB Accessibility
- EPUB/A has to align with all (mostly politics and textual niceties)

ISO Process

- Aug 2024 Draft submitted (ISO/IEC DIS 22424)
- Jun 2025 Comments received
- Aug 2025 Revisions and responses submitted
- Will replace ISO/IEC TS 22424-1:2020 (rather woolly) with a “harder” standard



Any Questions?

JH^{OVE}

veraPDF

fido

jpylyzer

ViPER

powered by veraPDF

OPF
TOOLS



eosc | EDEN
Enhancing Digital Preservation Strategies at European and National Level



ACTIVE
PROJECTS



ePUB

PERSONAL
PROJECTS

