

This is a test document. It uses the Web Capture feature.


Creative Commons CC0: Public Domain Dedication. To the extent possible under law, Johan van der Knijff has waived all copyright and related or neighboring rights to this work.

A community hub for digital preservation

**Join the Preservation Army**

**The OPF is fully supporting this November's "Just Solve the Problem",  
get involved now!**

**Featured Member**



Bibliotheque nationale de France  
Charter Member

**Identifying ebooks (File ID Hackathon)**

By [willp-bl](#) on 19 November 2012 - 3:53pm

Several of us at The British Library took part in the [CURATEcamp file id hackathon](#) on Friday.

We decided that one issue we could make a useful impact on was identification of various ebook formats. eBooks are an important content type for the British Library, especially with the expected implementation of non-print legal deposit legislation next year. For a long list of formats look here:

[http://wiki.mobilerread.com/wiki/E-book\\_formats](http://wiki.mobilerread.com/wiki/E-book_formats)

[Read more](#) [willp-bl's blog](#) [1 comment](#) [Log](#)

**C3PO: a content profiling tool for preservation analysis**

By [peshkira](#) on 19 November 2012 - 11:03am

In the last months, I have been researching the problem of large-scale content profiling for preservation analysis. I do this for a number of reasons. For one, I support the opinion that formats are just another property. Undoubtedly, a very important one, but knowing which formats you have is not sufficient for good [preservation planning](#) and actions.

[Read more](#) [peshkira's blog](#) [Log in](#) or [register](#) to post comments

**UPCOMING OPF EVENTS**

**WEBINAR** 'Working Collaboratively in Digital Preservation' 30 November

**SCAPE Training** 'Keeping Control, Scalable Preservation Environments for Identification and Characterisation' 6-7 December, Guimaraes, Portugal Registration closes 29 November

**Bloggers**

Identifying ebooks (File ID Hackathon)  
More from [willp-bl...](#)

---

C3PO: a content profiling tool for preservation analysis  
More from [peshkira...](#)

---

Depositing Data from Facebook to MediaWiki  
More from [tonisant...](#)

---

Watching Stars  
More from [pixelatedpete...](#)

---

Experimentation, simulation and prediction in DP: Emerging challenges  
More from [cbecker...](#)

**Recent comments**

File and mime types 3 days 21 hours ago

---

The preservation underground? 4 days 2 hours ago

---

Thanks to all participants for a great workshop! 1 month 2 weeks ago

---

This is excellent stuff... 2 months 1 week ago

---

Replies: rendering vs interacting 2 months 2 weeks ago

---

iPRES 2012 Presentation 3 months 2 days ago

---

Deadline extended: Contribute your research challenge! 3 months 5 days ago

---

Publishing Test Corpus 3 months 2 weeks ago

---

Careful normalisation... 3 months 3 weeks ago

---

FITS' only re-invention... 3 months 3 weeks ago

**Planet DigiPres**

Have some JHOVE with your turkey

---

NDSA Levels of Digital Preservation: Release Candidate One

---

Preserving Digital Archaeological Data

---

Identifying ebooks (File ID Hackathon)

---

C3PO: a content profiling tool for preservation analysis

---

Preliminary JHOVE fix list

---

Worldwide file ID hackathon

---

Preserving.exe: A Short List of Readings on Software Preservation

---

Depositing Data from Facebook to MediaWiki

---

Expanding JHOVE

[More](#)

**OPF Wiki Activity Feed**

SO18 Comparing two web page versions for web archiving

---

Pagealyzer

---

Pagealyzer

---

SO18 Comparing two web page versions for web archiving

---

Pagealyzer > installation manual-1-pagealyzer.pdf

[More](#)

