

# Ebooks: What do we care (for)?

Peter Cliff, Digital Preservation Team  
[digitalpreservationteam2@bl.uk](mailto:digitalpreservationteam2@bl.uk)

Frodo came in soon  
ing in the dark, deer  
asked.  
'Yes,' answered G

and then in a  
himself, he sang soft  
The Road goes ever on  
Down from the door wh  
Now far ahead the Road  
And I must follow, if I can  
Pursuing it with eager feet,  
Until it joins some larger way  
Where many paths and errand  
And whither then? I cannot say  
He paused, silent for a moment. Th  
without another word he turne  
away from the lights and voices in  
field and tents, and followed  
three companions went  
his garden, and tro  
sloping path  
place in  
too

# overview

Introduction	13.30 – 13.40
Hands-on with the devices	13.40 – 14.00
Brainstorm Properties	14.00 – 14.15
Properties Revisited	14.15 – 14.30
Break	14.30 – 14.45
Properties and business needs	14.45 – 15.15
Report Back	15.15 – 15.30
Summary and End	15.30 – 16.00 (probably earlier)

# the problem

“Digital preservation is the series of management policies and activities necessary to ensure the enduring usability, authenticity, discoverability, and accessibility of content over the long term.”

[http://www.portico.org/digital-preservation/wp-content/uploads/2011/10/fea\\_kirchhoff.pdf](http://www.portico.org/digital-preservation/wp-content/uploads/2011/10/fea_kirchhoff.pdf)

What do we need to look out for with ebooks?

What is an ebook anyway?

Migration is a common solution for digital preservation, but if we migrate what do we check to ensure the migration worked?

**What do we care for?**

# properties & communities

- What to preserve?
  - “essential attributes of a digital object which affect its appearance, behaviour, quality and usability...[that] must be preserved over time for the object to remain accessible and meaningful.”  
<http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008sigprops>
- Common properties
- Subjective properties
  - Properties are defined by communities of stakeholders. You'll probably read about the “designated community”.

# **session one: defining properties**

the sky  
 was can dy  
 lu nous mi ed  
 i  
 ble  
 spry pinks  
 shy lem  
 ons  
 greens  
 cool  
 choco lates  
 un der  
 a lo  
 co  
 mo tive s pout  
 ing  
 vi  
 o lets

# layout

l(a

le  
af  
fa

ll

s)  
one  
l

iness

s(

these out of in  
finite no  
where,who;arrive s  
trollingly

:alight whitely and.

)now  
flakes:are;guests,of t  
wi  
ligh

t

# structure

The screenshot shows the BaseX 7.7 database interface. The title bar reads "BaseX 7.7 - content". The menu bar includes "Database", "Editor", "View", "Nodes", "Options", and "Help". The toolbar contains various icons for file operations, navigation, and editing. The search bar on the right shows "1 Results".

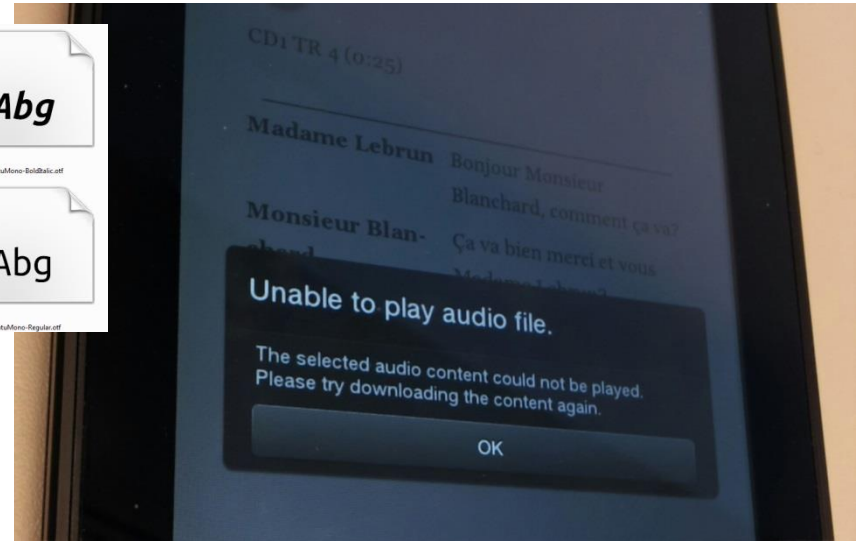
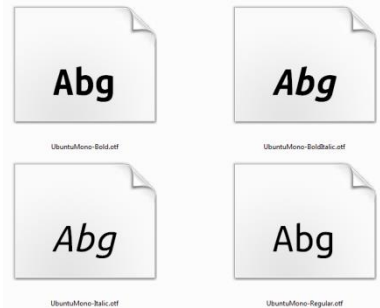
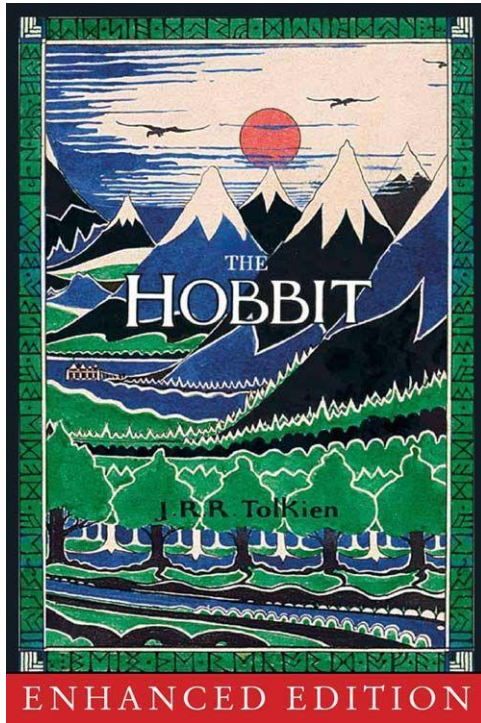
The main display area shows a hierarchical tree structure of the XML document. The root node is "content.opf", which contains a "package" node. The "package" node has four children: "metadata", "manifest", "spine", and "guide". The "manifest" node is expanded, showing a list of "text" elements. Below the tree, there is a row of 20 "text" elements, each containing a snippet of text from the XML document.

The bottom panel is divided into two sections: "Result" and "Query Info". The "Result" section displays the XML document structure in a tree view, with the "package" node selected. The "Query Info" section shows the command used to execute the query: "SET PARSEOPT encoding=UTF-8,flat=false,format=verbose,header=false,json\_spec=RFC4627,json\_unesc=true,jsonxml=false,lines=true,separator=comma". It also shows the command to set the language: "SET LANGUAGE English". The "Result" section shows the command to create the database: "CREATE DB content 'C:/EbookWorkshop/BooksForEach/ORF\_EPUB 3 Best Practices/OEBPS/content.opf'". The "Result" section shows the command to open the database: "db:open('content','content.opf')/package/manifest".

The status bar at the bottom right indicates "40 MB".



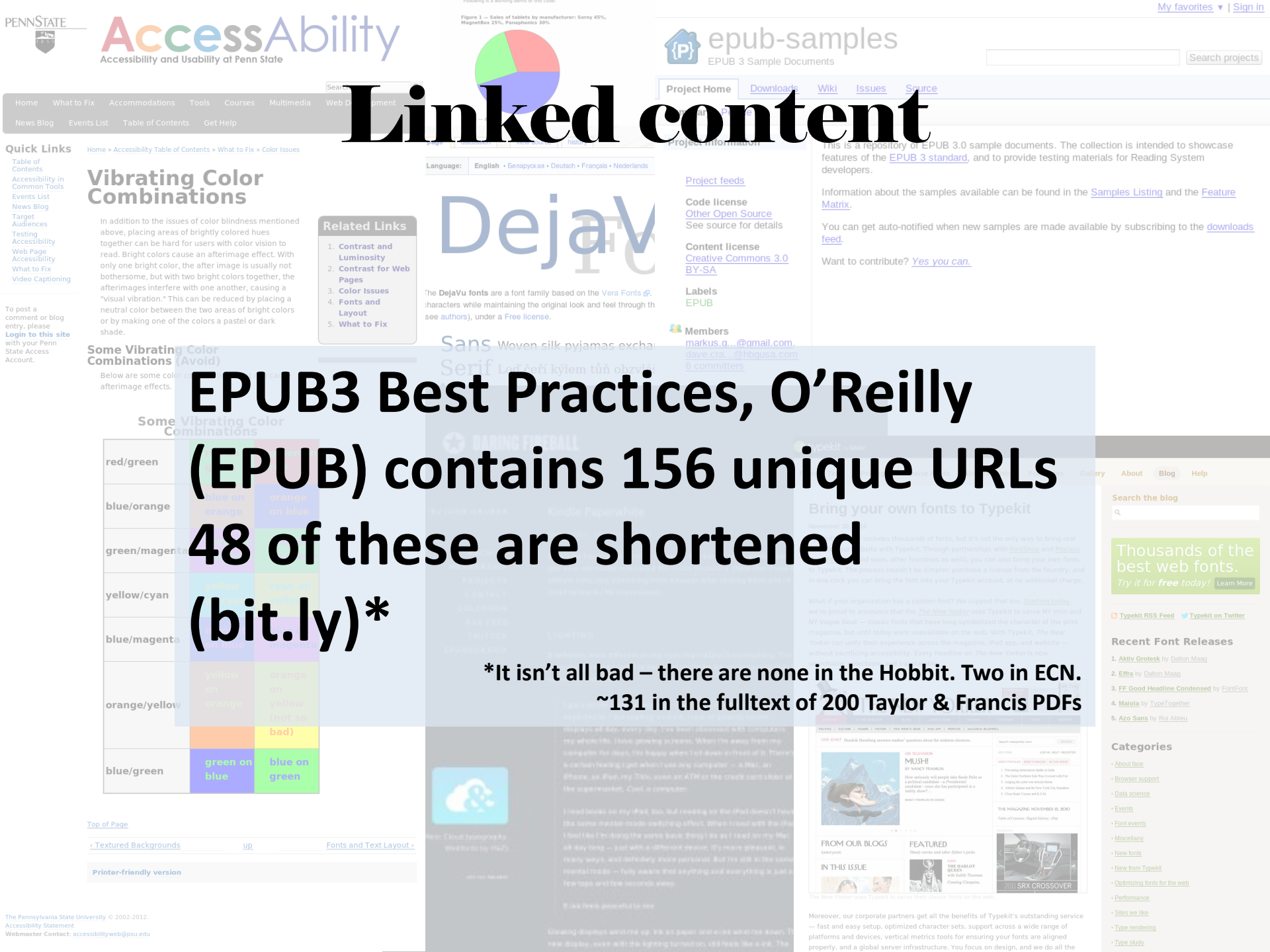
# embedded content



- Cover Images
- Inline Images
- Fonts
  - Obfuscation
  - Subsets
  - Licensing
- Audio
- Video

and anything else you can think of!





**Epub3 Best Practices, O'Reilly (Epub) contains 156 unique URLs 48 of these are shortened (bit.ly)\***

**\*It isn't all bad – there are none in the Hobbit. Two in ECN. ~131 in the fulltext of 200 Taylor & Francis PDFs**

# digital rights management

- Items linked to accounts
  - Will Adobe, Amazon & Apple systems last forever?
- Items that are password protected
- Items that have policies that restrict:
  - Printing
  - Cut and Paste
  - More than one user
  - More than one device

## **More Good News™:**

Ebook pilot books seem mostly open and the BL *may* have legal support in saying “No!” to DRM content.

# reading devices & software



Do we really want to preserve these too?

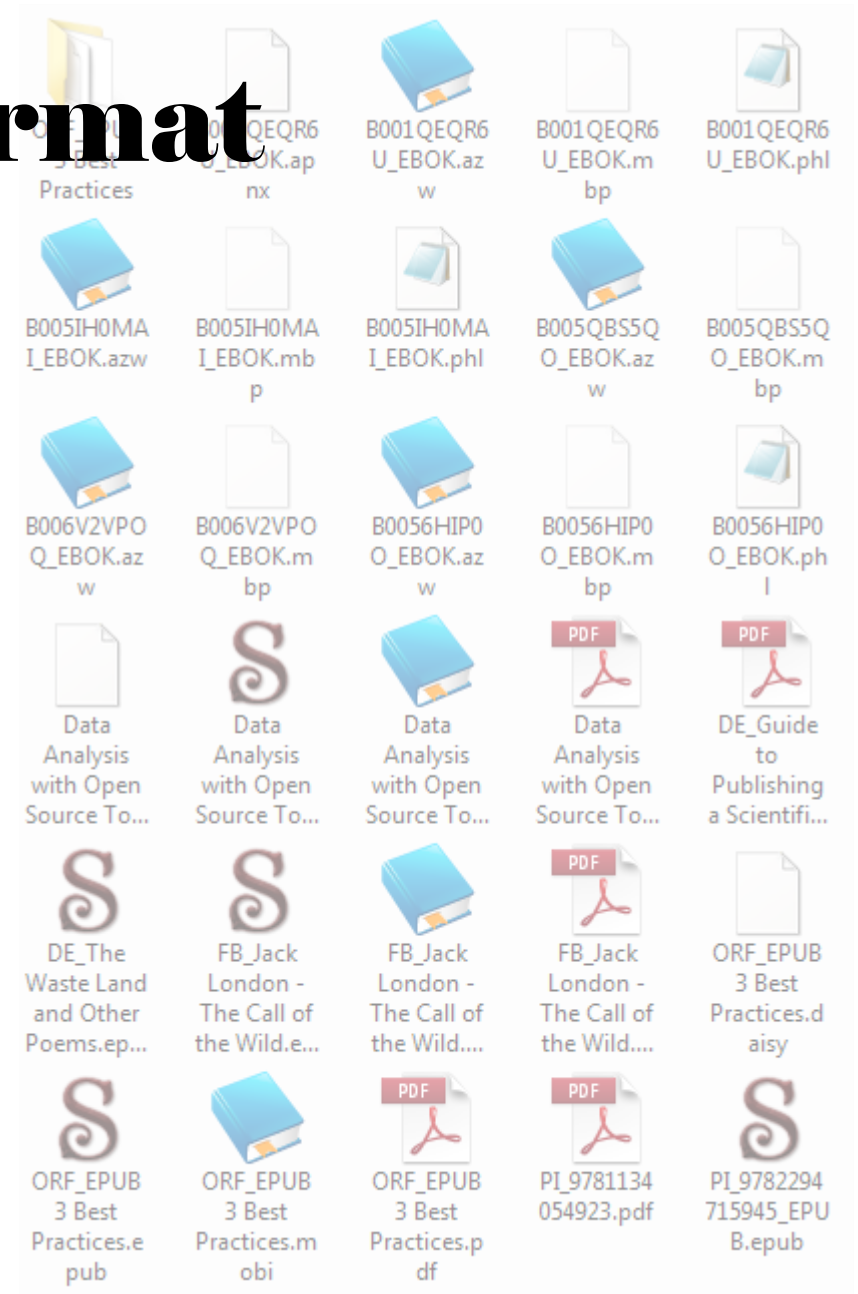
Is it the British Library's remit to?

(most batteries are not easy to replace and unlikely to stand the test of time)

DRM and devices are often linked

# file format

- Re-flowable formats
  - EPUB
  - MOBI
  - etc.
- Fixed layout formats
  - PDF
  - EPUB
- Odd ones
  - doc anyone?



# metadata & annotation

```
<popular update_id="10" timestamp="2013-07-28 23:33:00">
  <content version="0" guid=":8192B2FF" key="B005IH0MAI" type="EBOK">
    <annotation type="popular_highlight" num_users="4" end_pos="182191" start_pos="181927"/>
    <annotation type="popular_highlight" num_users="6" end_pos="230827" start_pos="230510"/>
    <annotation type="popular_highlight" num_users="6" end_pos="232850" start_pos="231655"/>
    <annotation type="popular_highlight" num_users="6" end_pos="233993" start_pos="232548"/>
    <annotation type="popular_highlight" num_users="6" end_pos="233993" start_pos="233449"/>
    <annotation type="popular_highlight" num_users="5" end_pos="235570" start_pos="234999"/>
    <annotation type="popular_highlight" num_users="5" end_pos="236118" start_pos="235971"/>
    <annotation type="popular_highlight" num_users="5" end_pos="237836" start_pos="237511"/>
    <annotation type="popular_highlight" num_users="5" end_pos="241095" start_pos="240549"/>
    <annotation type="popular_highlight" num_users="4" end_pos="706256" start_pos="706164"/>
  </content>
</popular>
```

```
<\bобнїтєк|
<\cоулєул>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="4" єулq_бoє="10є52є" єлєтєл_бoє="10єтєє"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="2" єулq_бoє="54т0є2" єлєтєл_бoє="540248"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="2" єулq_бoє="53183є" єлєтєл_бoє="5312тт"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="2" єулq_бoє="53єтт8" єлєтєл_бoє="53281т"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="2" єулq_бoє="532210" єлєтєл_бoє="534888"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="є" єулq_бoє="533883" єлєтєл_бoє="533448"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="є" єулq_бoє="535820" єлєтєл_бoє="535248"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="2" єулq_бoє="53т812" єлєтєл_бoє="53тє22"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="є" єулq_бoє="530851" єлєтєл_бoє="5302т0"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="є" єулq_бoє="т85т8т" єлєтєл_бoє="т8т851"\\>
<гунвогєгтєу cлбє="бобнїтєк_мїдмїтїдмє" унлв_лєєтє="є" єулq_бoє="т85т8т" єлєтєл_бoє="т8т851"\\>
```

- Bibliographic
- Technical
- Levels
  - E.g. chapter-level metadata
- Annotation



# versions

# T.S. Eliot The Waste Land and other poems

1. <http://www.gutenberg.org/ebooks/1321>
2. Eliot, Thomas Stearns. *The Waste Land*. New York: Horace Liveright, 1922
3. <http://www.bartleby.com/201/1.html>
4. <https://itunes.apple.com/us/app/the-waste-land/id427434046?mt=8>
5. <http://www.touchpress.com/titles/thewasteland/>
6. Eliot, Thomas Stearns. *The Waste Land*. Richmond, Surrey: Leonard & Virginia Woolf at the Hogarth Press, 1923
7. Eliot, Thomas Stearns. *The Waste Land*. London: Faber & Faber 1961
8. [http://www.amazon.co.uk/Waste-Land-Other-Poems-ebook/dp/B0056HIP00/ref=sr\\_1\\_1?ie=UTF8&qid=1377810829&sr=8-1&keywords=the+wasteland](http://www.amazon.co.uk/Waste-Land-Other-Poems-ebook/dp/B0056HIP00/ref=sr_1_1?ie=UTF8&qid=1377810829&sr=8-1&keywords=the+wasteland) [Kindle Edition]
9. <http://www.waterstones.com/waterstonesweb/products/t-s-eliot/the+waste+land+and+other+poems+28ebook29/8628472/> [EPUB DRM]
10. <http://www.waterstones.com/waterstonesweb/products/t-s-eliot/the+waste+land+28ebook29/6987513/> [EPUB DRM]

“Managing [the] complexity [of multiple versions of each manifestation] will be one of the unique challenges of eBook preservation”



# **session two: priorities**

# Summary

# Credits

Several quotes from:

[http://www.portico.org/digital-preservation/wp-content/uploads/2011/10/fea\\_kirchhoff.pdf](http://www.portico.org/digital-preservation/wp-content/uploads/2011/10/fea_kirchhoff.pdf)

<http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008sigprops>

Layout slide background taken from E.E.Cummings Selected Poems, W. W. Norton & Co. (Liveright), New York 1994

Structure Slide is using BaseX tree visualisation of the content.opf taken from O'Reilly's EPUB 3 Best Practices  
[<http://shop.oreilly.com/product/0636920024897.do>]

Embedded Content uses cover images from EPUB3 Best Practices and the Kindle Enhanced Edition of the Hobbit  
[<http://shop.oreilly.com/product/0636920024897.do> and [http://www.amazon.co.uk/The-Hobbit-Enhanced-Edition-ebook/dp/B005IH0MAI/ref=tmm\\_kin\\_title\\_2?ie=UTF8&qid=1378220173&sr=8-1](http://www.amazon.co.uk/The-Hobbit-Enhanced-Edition-ebook/dp/B005IH0MAI/ref=tmm_kin_title_2?ie=UTF8&qid=1378220173&sr=8-1)]

Versions slide uses cover image from The Waste Land and other poems, T.S.Eliot, Faber & Faber, London, 1972

Metadata and annotation slide background is a sample Amazon annotation file.

The Linked Content slide background shows sites extracted from the EPUB 3 Best Practices book and captured using cutycapt  
[<http://cutycapt.sourceforge.net/>].

Cover image shows an ebook reader and phone on the same page of The Hobbit and is by me.

So long folks!

# The End

He paused, silent for a moment. Then  
other he turned away from the  
ices in the field and followed  
three companions went round into his  
tted down the long sloping path. He  
place in the hedge at the bottom  
adows, passing into the night like  
the grass.

Gandalf remained for a while  
darkness. 'Good-bye, my  
xt meeting!' he said softly

Peter Cliff, Digital Preservation Team  
digitalpreservationteam2@bl.uk

Frodo came in soon  
ing in the dark, deer  
asked.

'Yes,' answered G

and then in a  
himself, he sang softly

The Road goes ever on  
Down from the door where  
Now far ahead the Road  
And I must follow, if I can  
Pursuing it with eager feet,  
Until it joins with larger ways  
Where many paths and errands  
And whither then? I cannot say

He paused, silent for a moment. Then  
without another word he turned  
away from the lights and voices in  
field and tents, and followed  
three companions went  
his garden, and took  
sloping path  
place in  
too