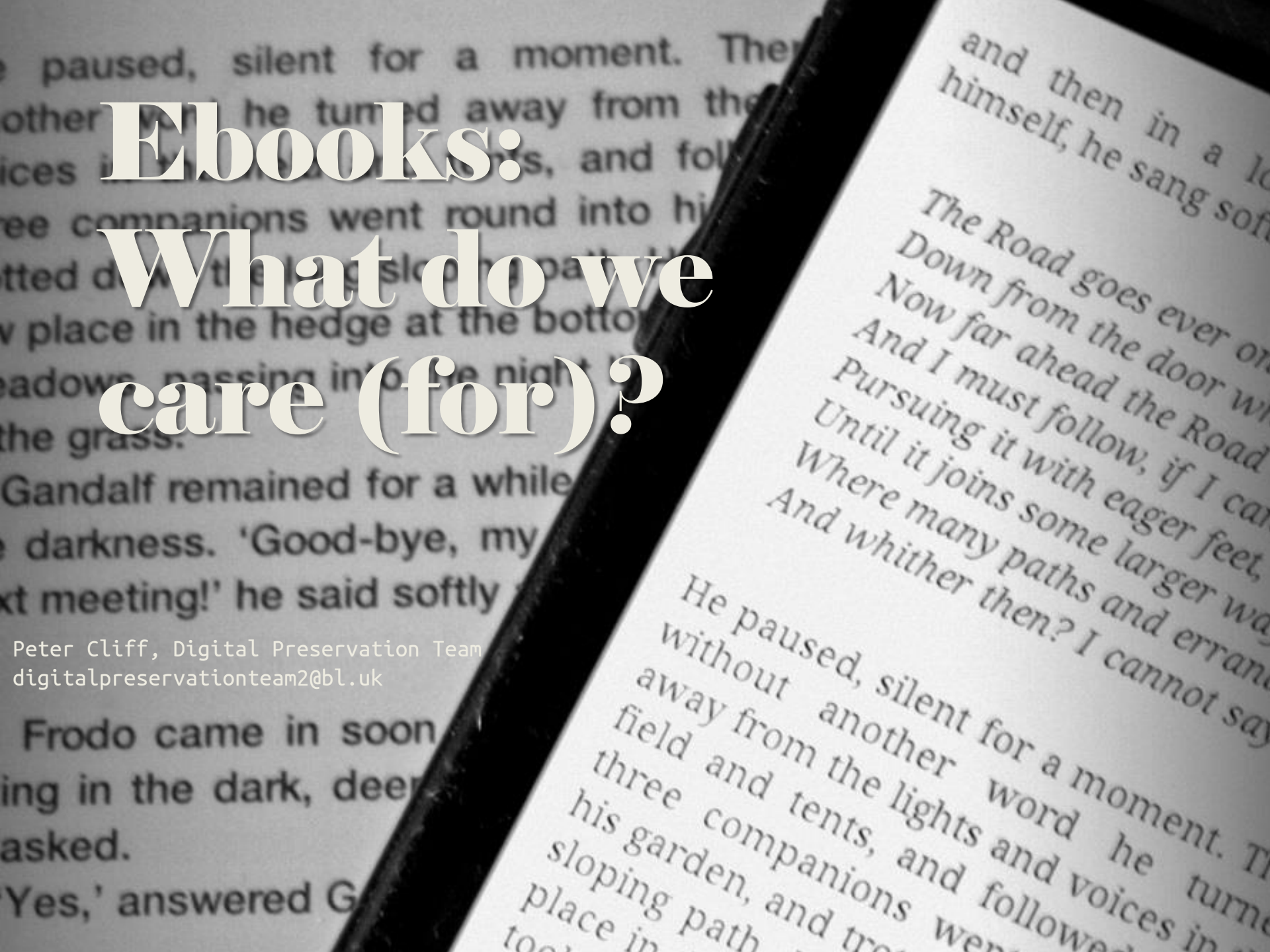
# Ebooks: What do we care (for)?

Peter Cliff, Digital Preservation Team
digitalpreservationteam2@bl.uk



# 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

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

# layout

## 3

the sky
was can dy
lu mi
nous 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



## 4

l(a

le
af
fa

ll

s)
one
l

iness

## 5

s(

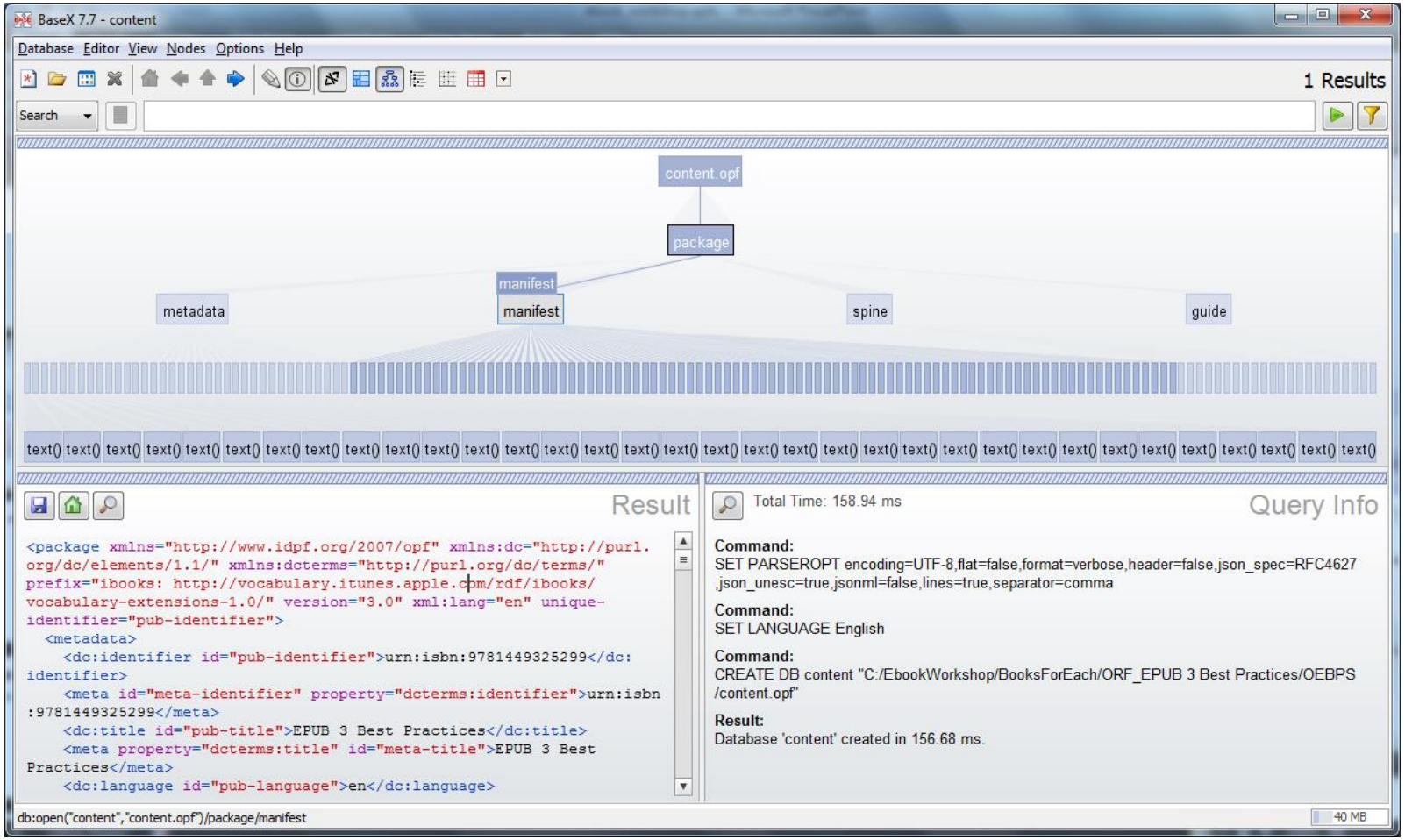
these out of in
finite no
where,who;arrive s
trollingly

:alight whitely and.

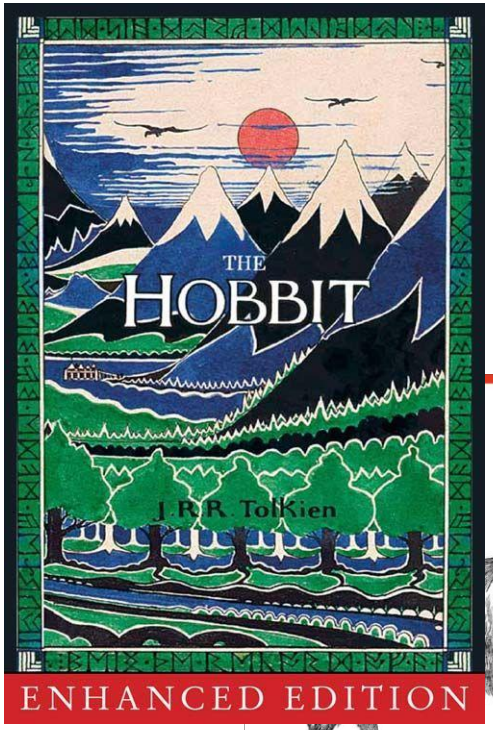
)now
flakes:are;guests,of t
wi
ligh

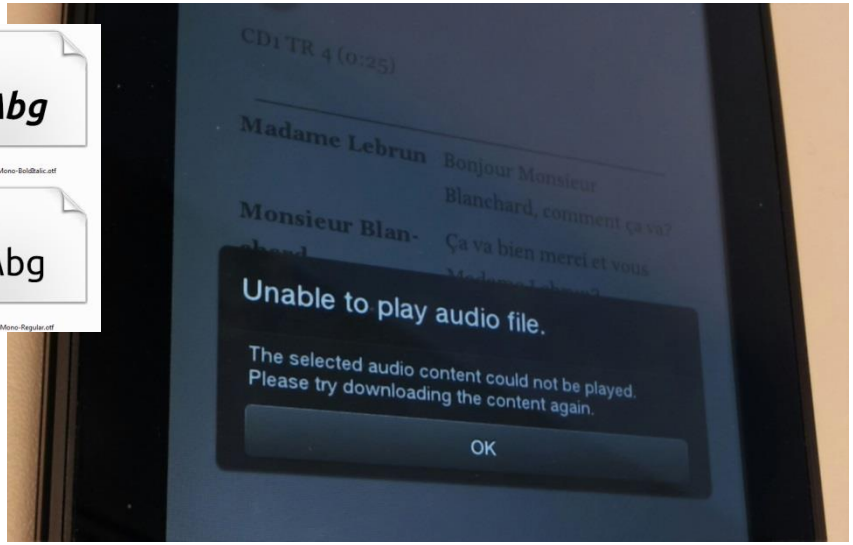
t

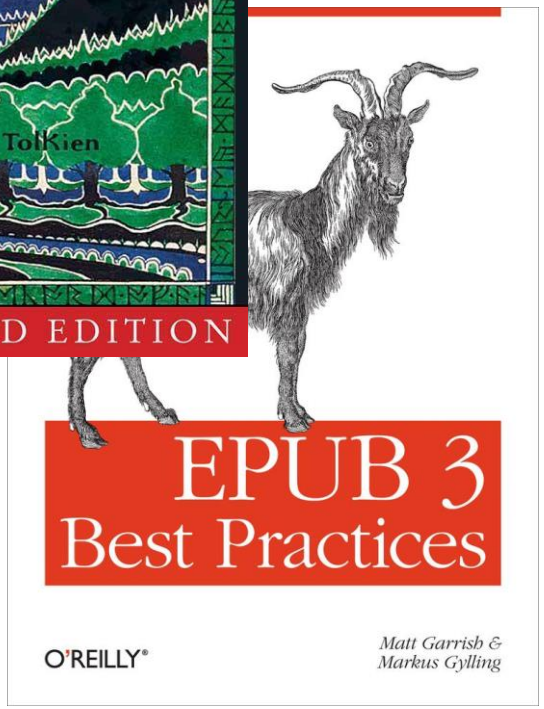
# structure


BaseX 7.7 - content
Database Editor View Nodes Options Help
1 Results
Search
content.opf
package
manifest
metadata
manifest
spine
guide
text() text() text()

# embedded content






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

and anything else you can think of!

# Linked content

## 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**

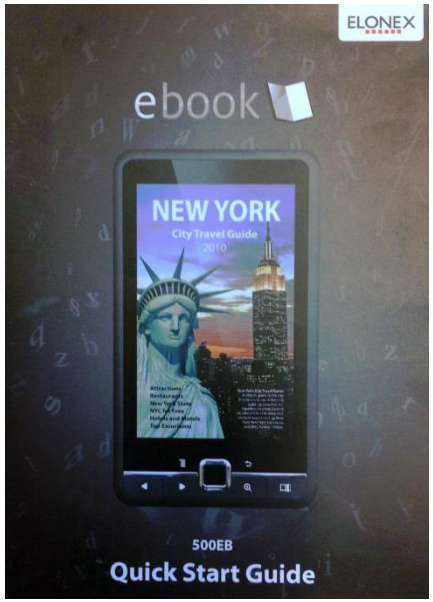
# digital rights management

- 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

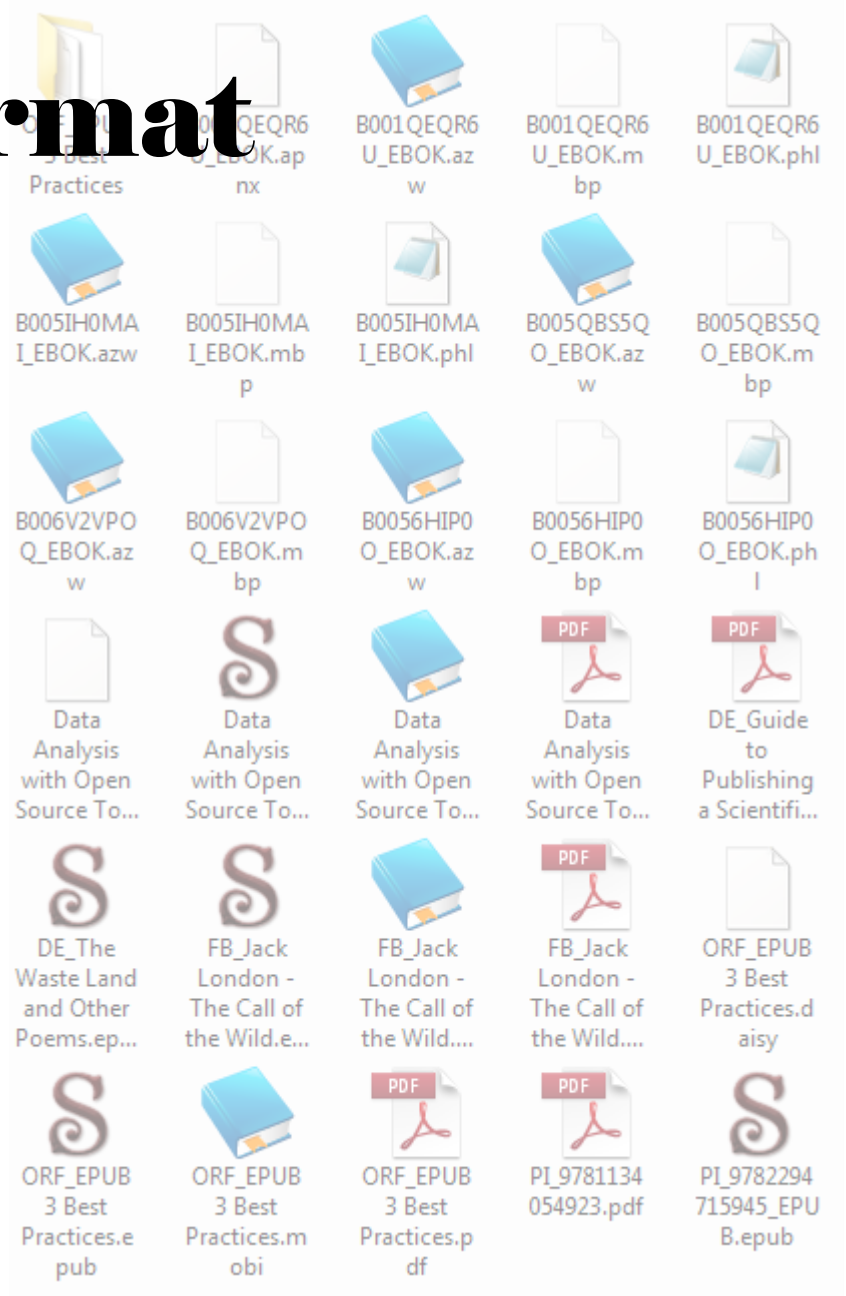



Do we really want to preserve these too?
Is it the British Library's remit to?
(most batteries are not easy to replace and unlikely to stand the test of time)

DRM and devices are often linked

# file format

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



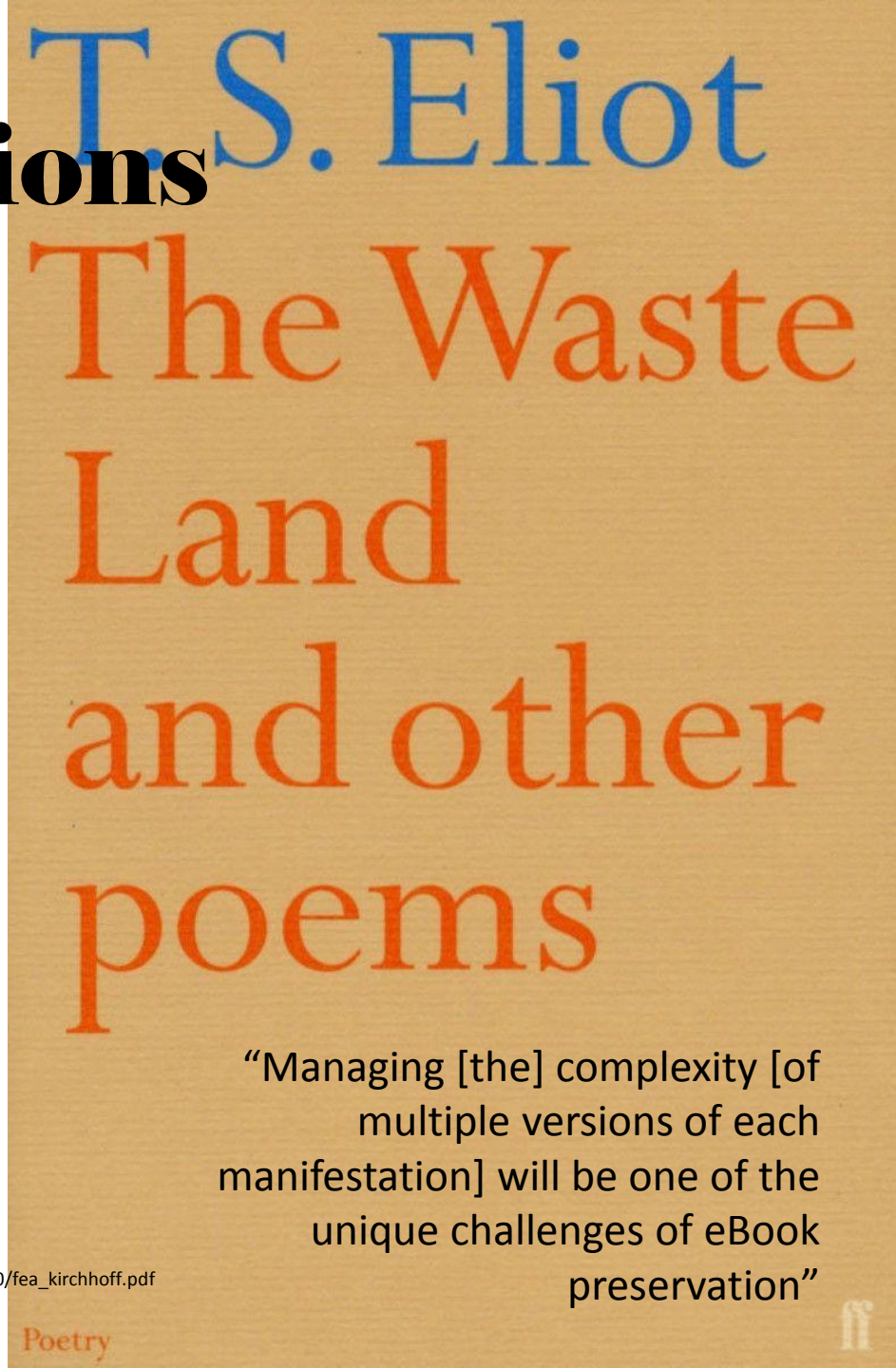
# metadata & annotation


```
<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="5" end_pos="231975" start_pos="231655"/>
    <annotation type="popular_highlight" num_users="6" end_pos="232850" 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>
```


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

# versions

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/B0056HIP0O/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"

http://www.portico.org/digital-preservation/wp-content/uploads/2011/10/fea_kirchhoff.pdf

# session two: priorities

# Summary

# Credits

Several quotes from:
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]

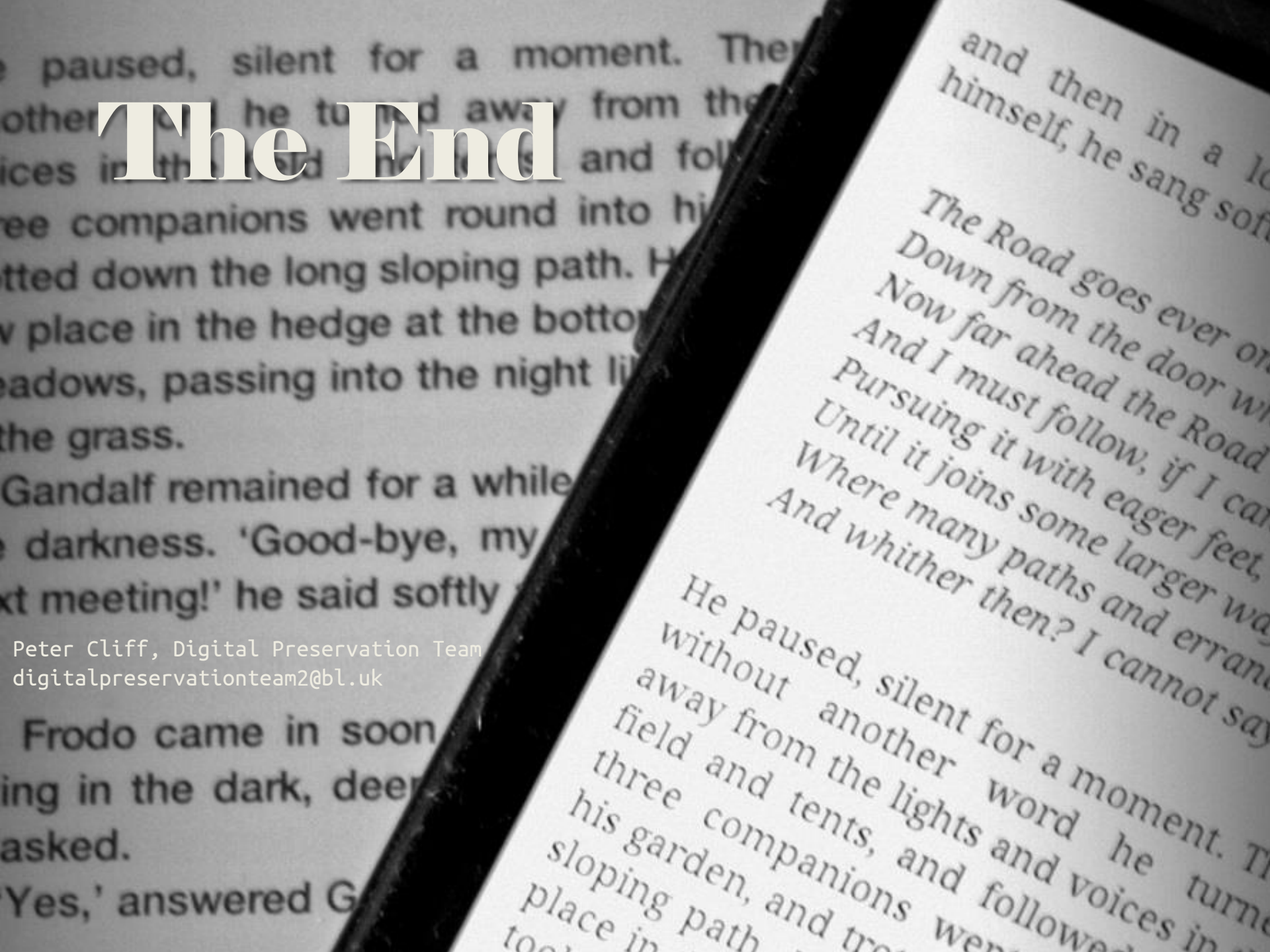
Versions slide uses cover image from The Waste Land and other poems, T.S.Eliot, Faber & Faber, London, 1972

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!




Peter Cliff, Digital Preservation Team
digitalpreservationteam2@bl.uk