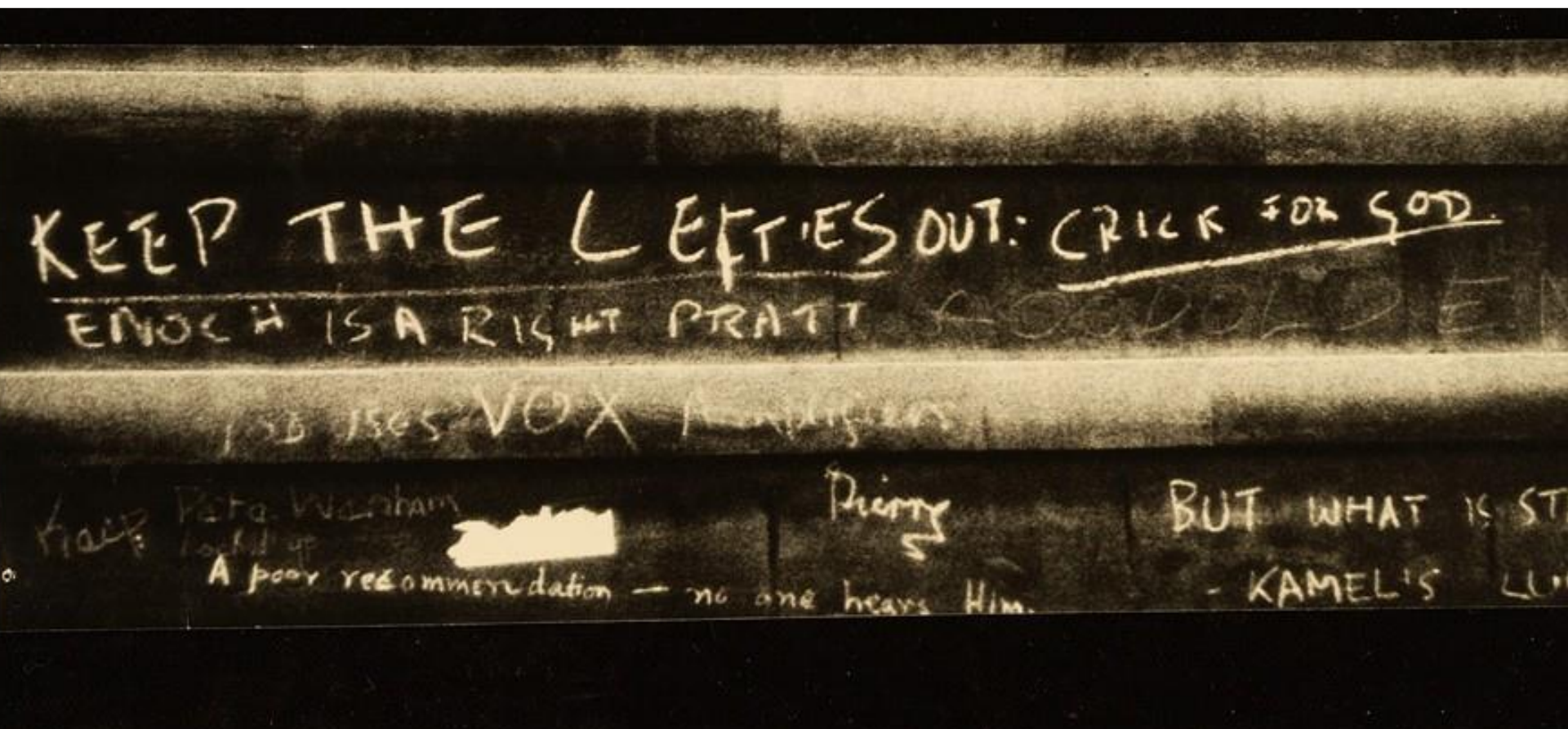
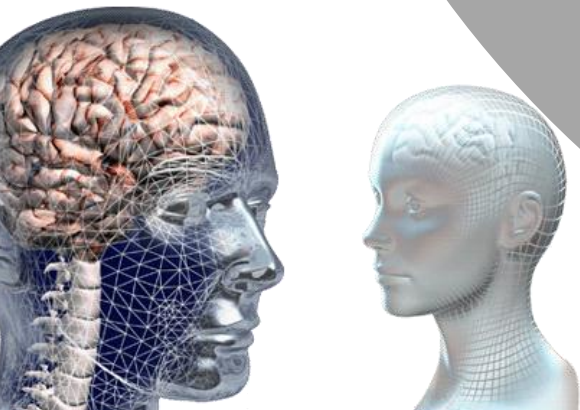


Digitising modern archives – the *Codebreakers* project



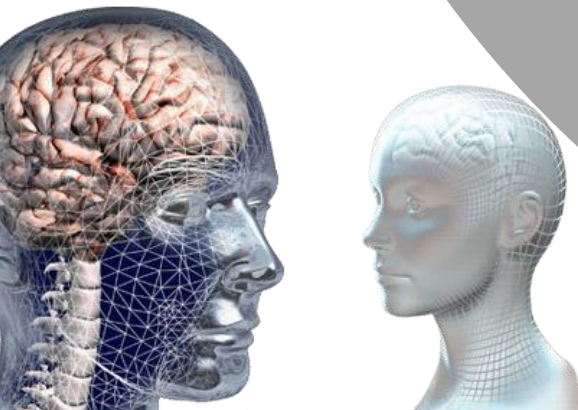
Wellcome Trust

A global charitable foundation
dedicated to achieving
extraordinary improvements
in health by supporting the
brightest minds.



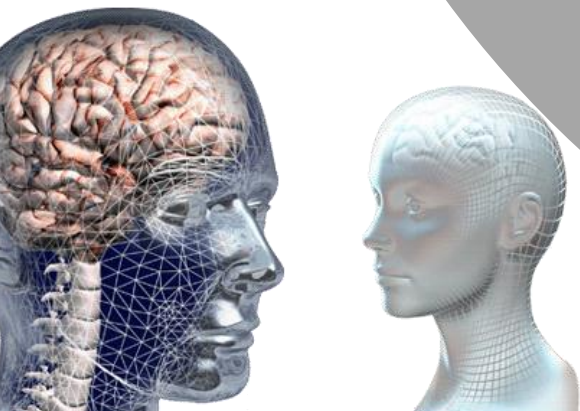
Wellcome Trust

We strive to embed
biomedical science in the
historical and cultural
landscape.



*{ the free library for
the incurably
curious }*

Wellcome Library

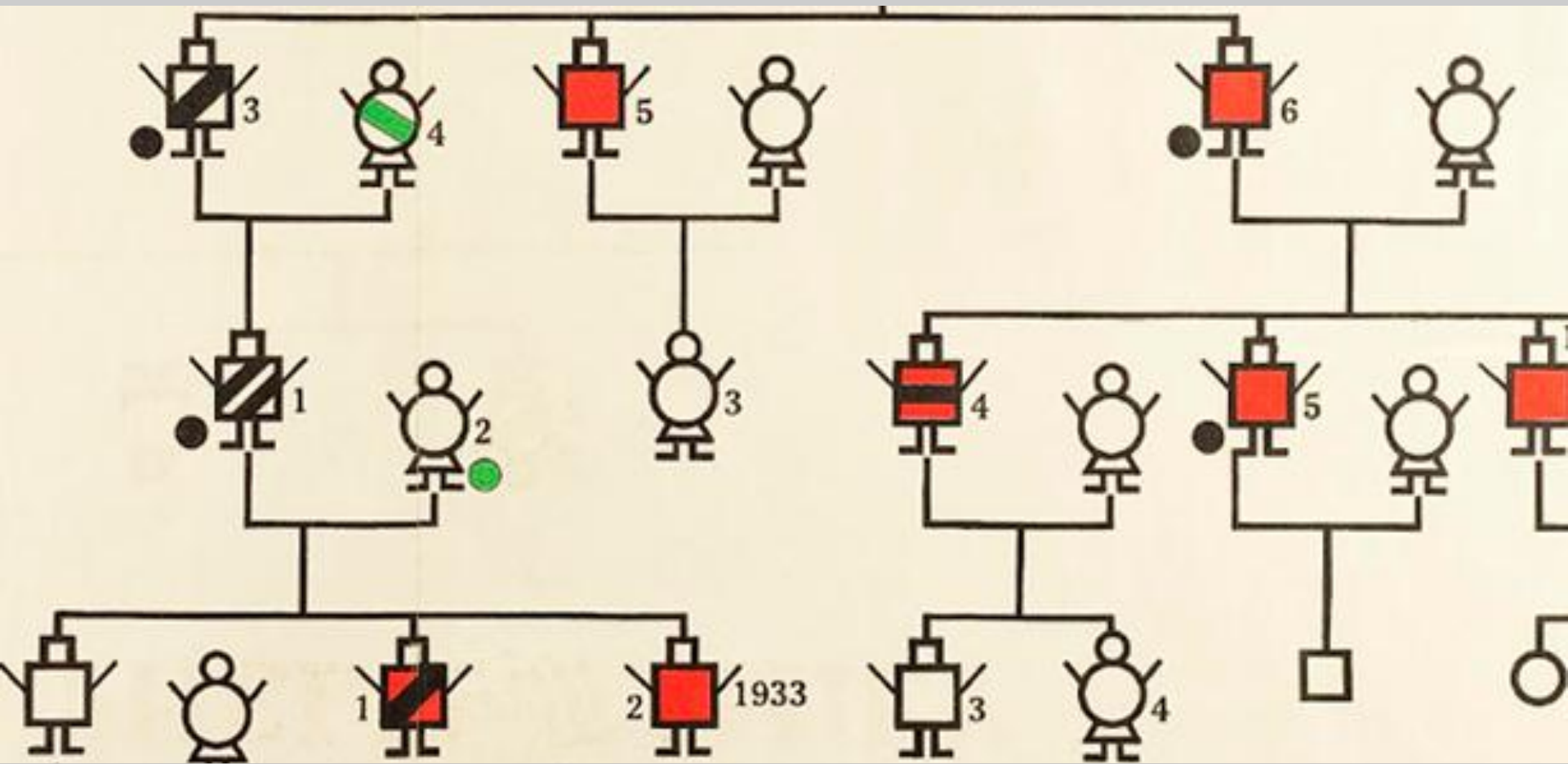


{ we are
transforming
ourselves }

The Wellcome Library aims to digitise 50 million pages of books and archives, to create a free online research library based on our holdings, and those of partner institutions.



{ *Codebreakers: makers of modern genetics* }



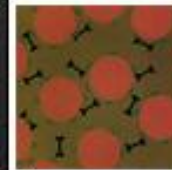
Our pilot project involved digitising 1,700 books and 22 archive collections drawn from six institutions – about 1.5m pages of content in total – together with interpretive content.

THE HISTORY OF GENETICS: FROM DARWIN TO THE 21ST CENTURY: 1859 - 2012

Not surprisingly, the content is heavily geared towards the 20th century. A focus on molecular biology shifted emphasis still further toward the present.



1859
CHARLES DARWIN
PUBLISHES 'ON THE
ORIGIN OF...



1901
KARL LANDSTEINER
DISCOVERS ABO
BLOOD GROUP...



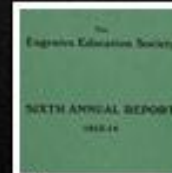
1866
GREGOR MENDEL
PUBLISHES HIS
THEORY OF...



1869
FRANCIS GALTON
PUBLISHES
'HEREDITARY
GENIUS'



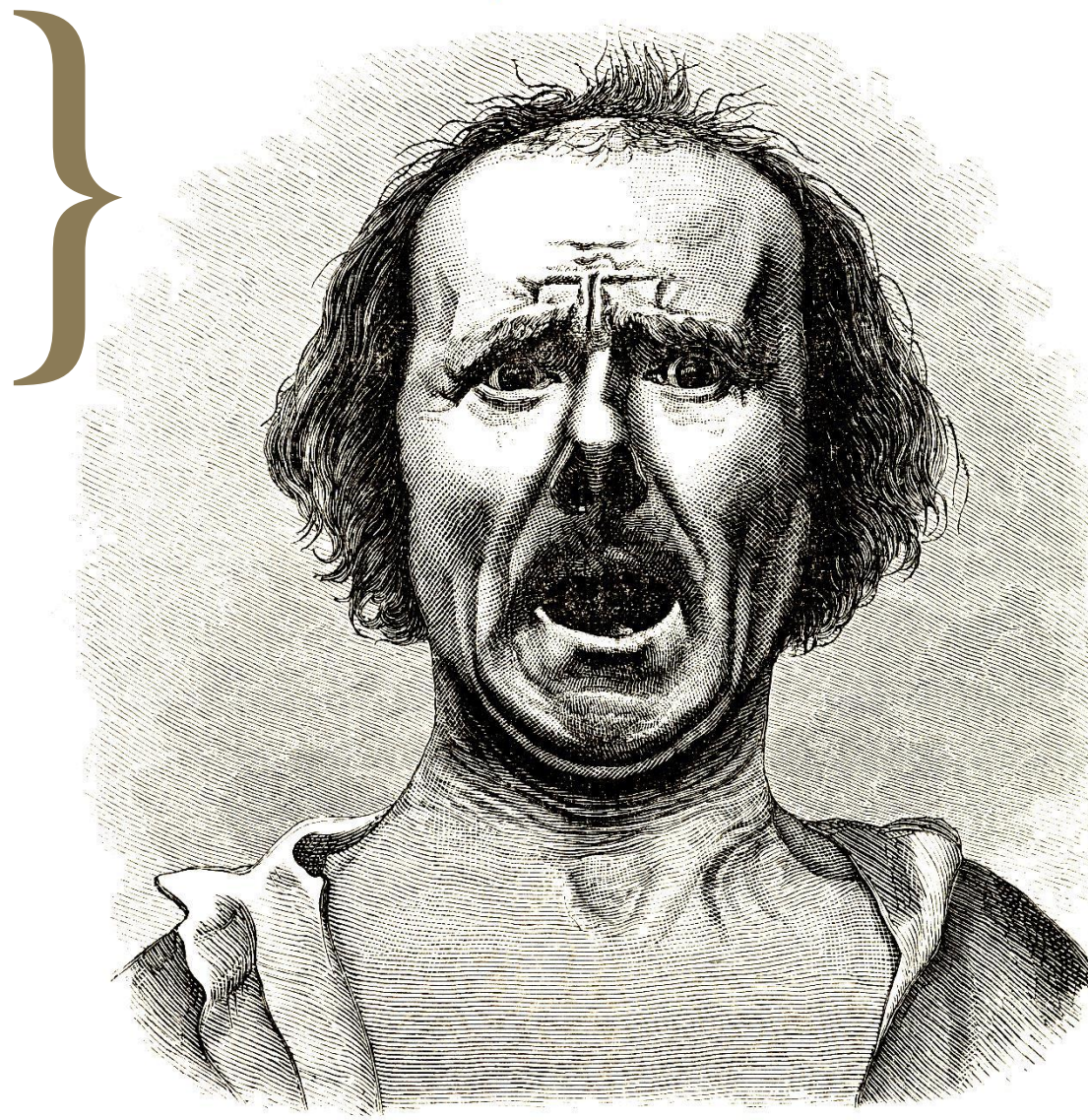
1909
TREASURY OF
HUMAN
INHERITANCE
LAUNCHED



1913
UK MENDEL
DEFICIENCY



but what about
sensitivity
and
copyright?



{copyright}

In the UK all published material is in copyright until 70 years after the authors death. Unpublished material is in copyright until 2040.

{sensitivity}

The Data Protection Act (1998) restricts access to data about living people. Archives also hold potentially sensitive data about physical/mental health and sexual lives, and genetic heritage.



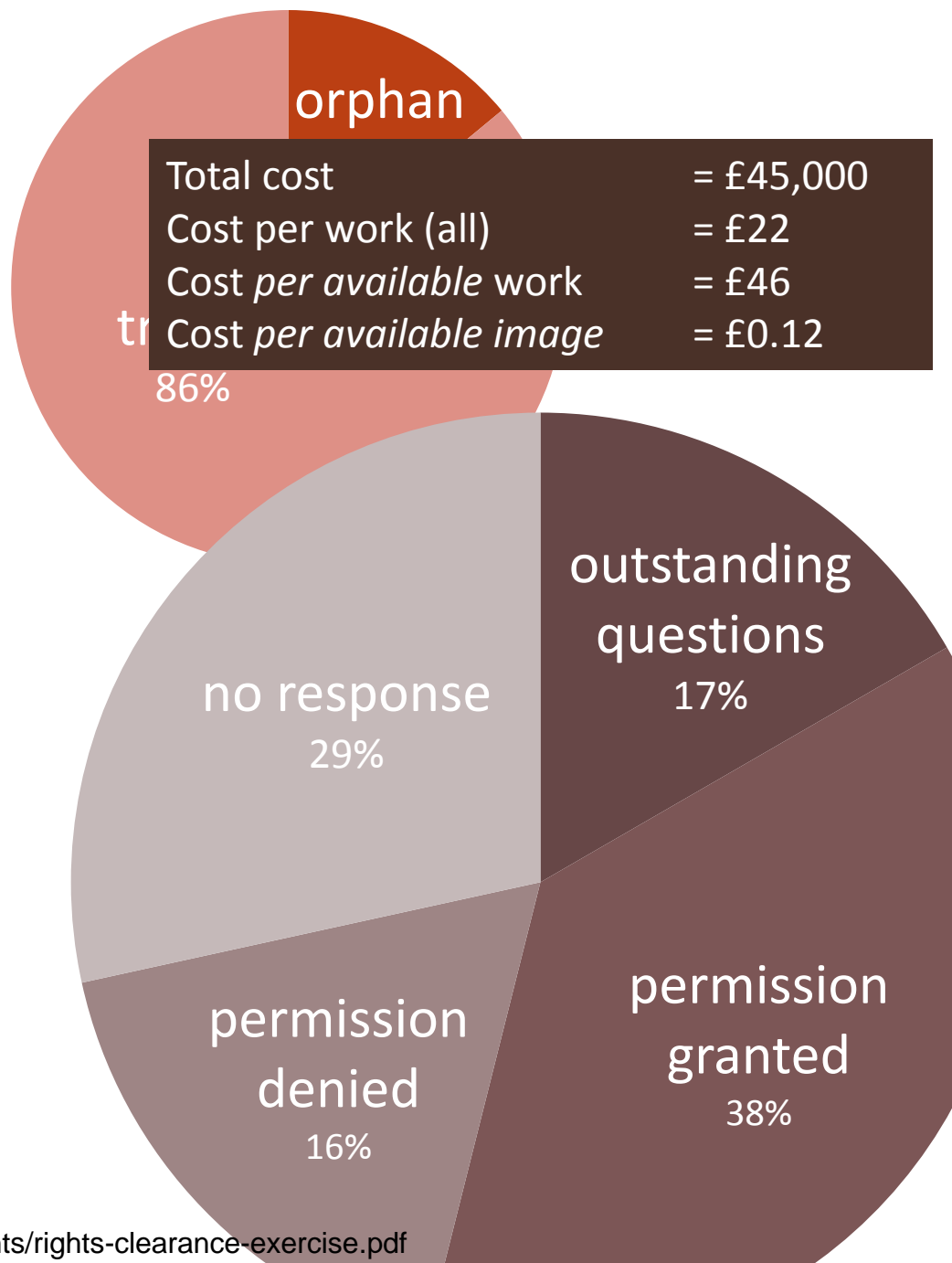
{ *case study:* copyright for books }

Worked with ALCS/PLS to identify rightsholders.

3/3 rule – for multi-author works we only looked at first 3 named authors, or authors named in 3 or more works.

We discovered 5,500 authors for 1,620 works - most had 1 or 2 authors, but 32 works had 41+ authors each.

Wrote (3 times) to each rightsholder.



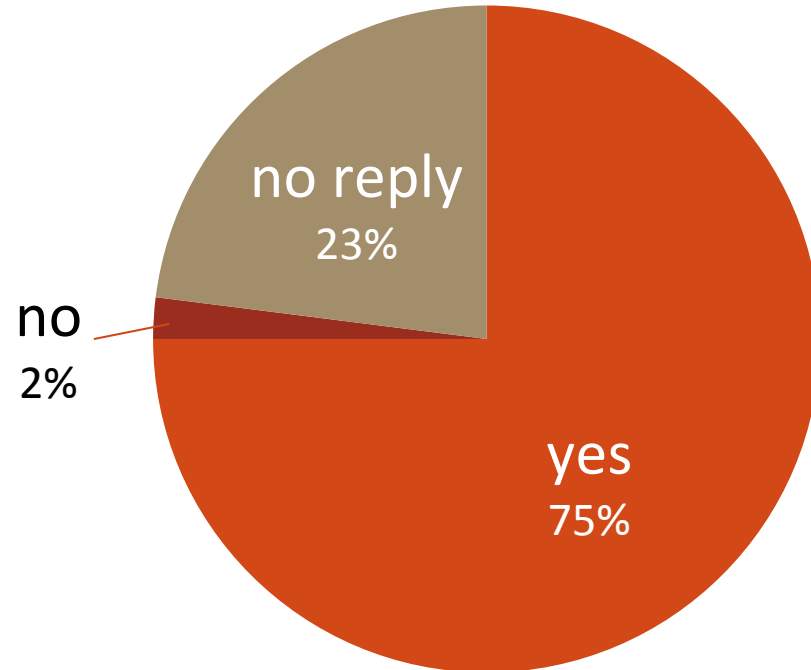
case study: **copyright for archives**

Very many copyright holders – and very few identified in catalogue metadata.

Unstructured data – one folder can contain hundreds of potential rights holders.

Risk-managed approach is essential – identify those who have most to lose.

Selective identification and contacting, emphasising the benefits of saying yes.



total cost = <£12,000

(distributed among partners)

cost per available image = £0.005 - £0.01

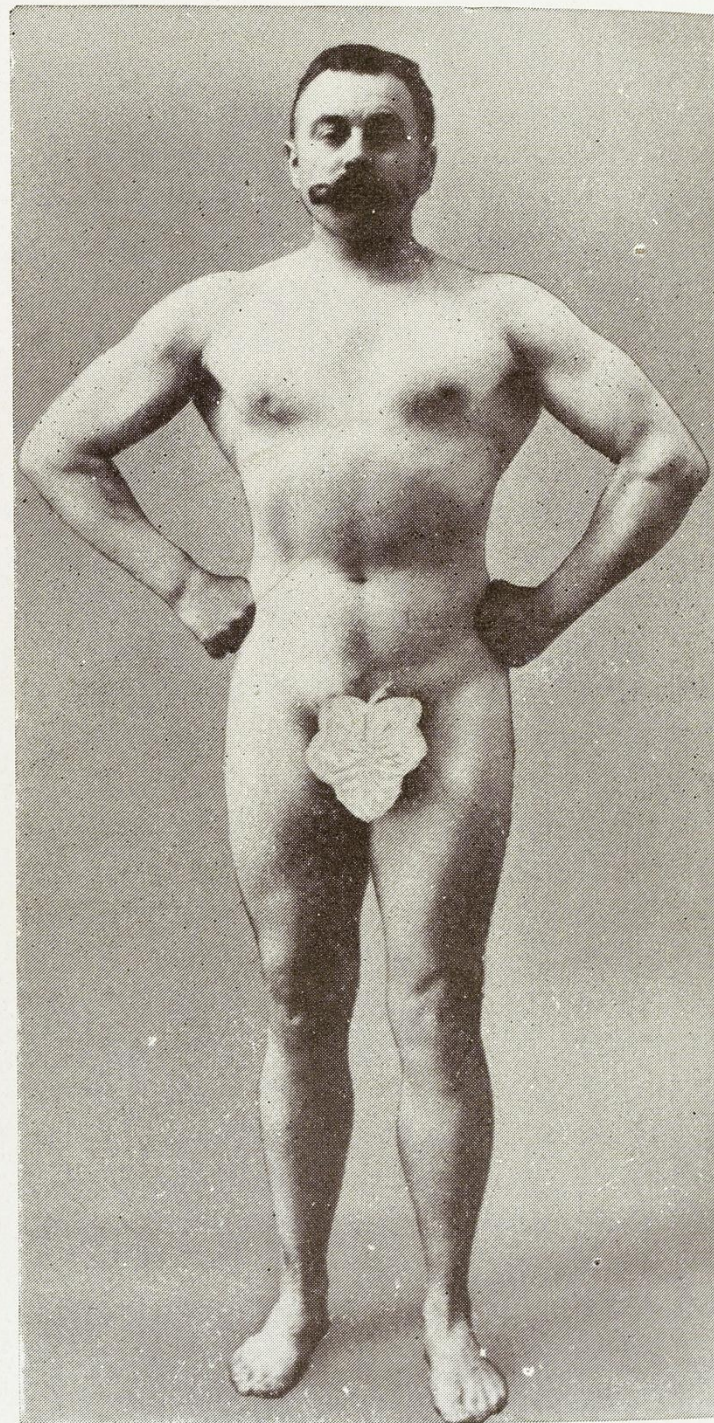
case study: **sensitivity for archives**

Legal obligation to abide by DPA – but best practice is to go beyond letter of law.

Can't check every item, instead sampled (and re-sampled where necessary). For new cataloguing, assessing sensitivity for digitisation can be tackled up front.

Not checking everything means risk exists – can be mitigated using access controls (but at a price).

Material classed as open, 'restricted' or closed. Anything <100 years old is restricted by default.



{ managing *access* }

Viewing 'restricted' requires login using library membership or social media (but this deters >95% of visitors, and we're changing to click-through T&Cs)

Also have ability to restrict download/embed, or restrict individual images within items.

Links to licence/credit info.

Clear access policy.

Robust takedown policy (and training to back it up)




In summary...

- ☐ You can't digitise modern archives without tackling sensitivity and copyright
- ☐ But unless you have VERY deep pockets, you can't be exhaustive
- ☐ Need to assess, and manage, the risk (to rightsholders/data subjects/you)
- ☐ For a collaborative project, need to ensure common approach
- ☐ Partners must ensure metadata is consistent, identify names, contact depositors
- ☐ Coordinator – consolidate and assess data, manage contacts with other rightsholders
- ☐ Building confidence may mean starting cautious, and loosening up
- ☐ You will need patience...
but it can be done





Further reading...



CREATE

RCUK Centre for Copyright and
New Business Models in the
Creative Economy



Home Team Research Programme Events Resources Policy Responses Publications Courses CREATE Blog Contact

CREATE » Publications » Copyright & Risk: Scoping the Wellcome Digital Library Project

Copyright & Risk: Scoping the Wellcome Digital Library Project

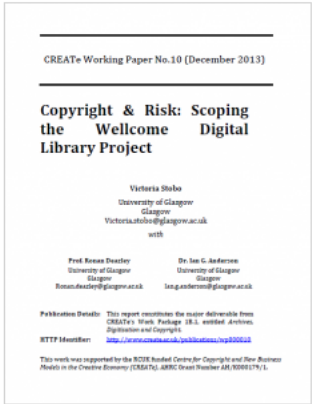
CREATE Working Paper No. 10

Victoria Stobo with Ronan Deazley and Ian G. Anderson (2013)


Introduction


Copyright & Risk: Scoping the Wellcome Digital Library is a comprehensive case study which aims to assess the merits of the risk-managed approach to copyright clearance adopted by the Wellcome Library (WL) in the course of their pilot digitisation project Codebreakers: Makers of Modern Genetics.


Whilst the WL clearly identifies as a library, as a collecting institution it extends from books and periodicals to include archives and manuscripts, art, the moving image and sound associated with the biomedical sciences and the medical humanities.[1] In this case study we consider the digitisation of both library (books) and archive material as part of the Codebreakers project, although our principal interest lies with the archive digitisation strand.



CREATE News

 **CREATE** @copyrightcentre 24 Apr
#filesharingreview
Uploaded event media - report pdf, txt transcript, audio, press bytes, social media, photos! bit.ly/1f6saEI
Expand

 **CREATE** @copyrightcentre 15 Apr
CREATE PhD Studentship in Cultural Heritage Law at University of Glasgow
bit.ly/1qY0EsA
Expand

 **CREATE** @copyrightcentre 11 Apr
Launch of File Sharing Review and panel
Tweet to @copyrightcentre

Subscribe to CREATE

Email Address*

<http://www.create.ac.uk/publications/copyright-risk-scoping-the-wellcome-digital-library-project/>

wellcomelibrary.org

wellcomeimages.org