CREATe Working Paper No.10 (December 2013)

Copyright & Risk: Scoping the Wellcome Digital Library Project

Victoria Stobo

University of Glasgow
Glasgow
Victoria.stobo@glasgow.ac.uk

with

Prof. Ronan Deazley
University of Glasgow
Glasgow
Ronan.deazley@glasgow.ac.uk

Dr. Ian G. Anderson
University of Glasgow
Glasgow
Ian.g.anderson@glasgow.ac.uk

Publication Details: This report constitutes the major deliverable from CREATe's Work Package 1B.1, entitled *Archives, Digitisation and Copyright*.

HTTP Identifier: http://www.create.ac.uk/publications/wp000010

This work was supported by the RCUK funded *Centre for Copyright and New Business Models in the Creative Economy (CREATe)*, AHRC Grant Number AH/K000179/1.

Acknowledgments

The staff of the Copyright and Risk: Scoping the Wellcome Digital Library Project would like to thank the Wellcome Library for their generosity in allowing CREATe access to the Codebreakers Project and the Wellcome Digital Library team.

Particular thanks to Sue Davies for organising the interviews at the Wellcome Library which form the basis of this report, and to those who gave up their time to speak to the RA: Simon Chaplin, Christy Henshaw, Caroline Herbert, Jenny Haynes, Toni Hardy, Chris Bird and Rada Vlatkovic; the information we gathered is much appreciated.

Especial thanks are due to Caroline Herbert for providing much of the data this report is based on; and also to Caroline (again), Christy and Sue for their comments on the draft.

Thanks also to the staff at the third party archives, contributors to Codebreakers and patient interview participants: Lesley Richmond and Sam Maddra at Glasgow University Archive Services; Allen Packwood, Natalie Adams and Andrew Riley at Churchill Archives Centre; Patricia Methven and Geoff Browell at King's College London Archives and Information Management; Mila Pollock and John Zarillo at Cold Spring Harbour Laboratory Library and Archives; and Gill Furlong, Katy Makin and Steve Wright at University College London Special Collections.

Table of Contents

Acknowledgments 2
1 Introduction 5
2: Introducing Codebreakers: Makers of Modern Genetics 8
2.1 The Wellcome Digital Library 8
2.2 Codebreakers: Makers of Modern Genetics 9
2.3 The Partner Archives 10
3 Archives Digitisation 13
3.1 Archives Digitisation at the WL 13
3.2 The WL Digitisation Workflow 14
4 Copyright 17
4.1 Copyright in Library Material 17
4.1.1 Library Clearance Results 17
4.1.2 Library material conclusion 22
4.2 Copyright in Archive Material 22
4.2.1 The Collaboration 24
4.2.2 Archive material clearance results 24
4.2.3 Post Project Lessons Learned from Clearing Copyright in Archive Material 31
5 Sensitive personal data 33
5.1 Sensitive personal data in context 33
5.2 Dealing with sensitive personal data at the WL 34
5.3. Dealing with sensitive personal data at the 3rd party archives 36
5.4 Conclusion on sensitivity checking 37
6 Conclusions 38
6.1 The success of the WL approach to clearing rights in archive material 38
6.2 The Policies 40
6.2.1 The Takedown Policy 40
6.2.2 Risk Management 40
6.2.3 Access to Archives 42
6.3 Reputation 42
6.4 Communication 43
6.5 Evidence to collect about archival rights clearance exercises 44
7 Appendixes 45
7.1 Example of an Archives Permission Letter 45
7.2 Example of a Library Permission Letter 50
Annex B: Copyright Clearance Form 53
7.3 WL Archive Risk Criteria 56

8 Bibliography..... 57

1 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.

This report assesses the merits of the WL's approach to copyright compliance by:

- Introducing the WL and the Codebreakers mass digitisation project
- Outlining the copyright challenges presented by the mass-digitisation effort
- Discussing the risk management methods used by the WL in determining which rights should be cleared
- Examining the results of the risk-managed copyright clearance process

As a result of this analysis, the research aims to:

- Compare the Wellcome experience on Codebreakers to rights clearance exercises attempted by other cultural heritage institutions
- Provide policy makers with useful data and insights to inform the debate on copyright exceptions for cultural heritage institutions
- Discuss the relevance of this approach for other UK archive institutions

Prior to the preparation of this case study, CREATe published a working paper in March 2013[2], reviewing current and proposed changes to UK copyright law, and specifically exceptions provided for libraries and archives. This was further extended by a review of available literature on the digitisation of archival and library collections for publication online, contextualized with specific examples.

Existing literature reveals that rights clearance procedures impose prohibitive burdens on cultural institutions, through the cost of staff time and training in both diligent search and the process of contacting rightsholders.[3] It also indicates that in most cases, the results of rights clearance processes are unsatisfactory: either copyright holders cannot be identified and traced; or those who are contacted, do not respond to permission requests. Archives, in contrast to

---

[1] The Wellcome Library Collection Development Policy is available at http://wellcomelibrary.org/content/documents/policy-documents/collection-development-policy.pdf [Accessed: 5 Dec 2013]

[2] Deazley, R., and Stobo, V. (2013) Archives and Copyright: Risk and Reform. Working Paper. CREATe / University of Glasgow, Glasgow, UK.

[3] *Ibid*

libraries, have the added complication of dealing with larger and more varied collections of material, the majority of which has been created for non-commercial purposes; this material is often unpublished at the point of deposit with the archive, and typically includes higher proportions of orphan works (when compared with traditional library collections).[4]

Documented examples of archival rights clearance projects are scarce, and where studies do exist, they have generally been conducted at a large-scale level, and do not contain sufficient detail to enable in-depth analysis of the rights clearance process.[5] Anna Vuopala's excellent Orphan Works study[6] includes examples of rights clearance projects at 19 institutions, and focuses almost exclusively on the time and cost required to clear rights, which is extremely useful, but leaves out other valuable details about the right clearance process: for example, the number of rightsholders' contact details found; the refusal rates and reasons given for refusal at each institution; and the number of rightsholders who do not respond to permission requests.

In general, there is a lack of detailed evidence concerning rights clearance in archival digitisation, and especially in relation to projects that employ risk-management strategies in a sector which is known to be highly risk averse. The WL was chosen as an appropriate subject for this case study given the scale of their digitisation aims, and their use of a risk managed approach to copyright compliance.

Codebreakers, as a digitisation project, was never designed to facilitate a study of the process of rights clearance - by the time data collection for the Copyright and Risk Project began, Codebreakers was entering its final stages, with the official launch of the project website taking place in February 2013. As a result, semi-structured interviews were chosen as the most effective method of obtaining data about the project. Interviews were conducted with key project staff at the WL as well as with staff from five partner archives involved in delivering the Codebreakers project.

During the interviews, and through participants' description of workflow processes, it became clear that checking material for sensitive personal data in the archive collections was of great concern to project staff. While the Copyright & Risk project team wished to look at sensitive personal data in greater depth, we were unable to do so within the bounds of the project; the report includes a short chapter explaining the approach to sensitive personal data at the WL and other archives. Added to comprehensive rights clearance, sensitivity checking is another huge burden to mass digitisation projects, and like archive rights clearance, very little has been written in this area, and further research is needed.[7]

The semi-structured format of the interviews was chosen for its flexibility; a list of questions was chosen, but given the length of the interviews, it was possible to divert from these questions and follow up on unexpected lines of enquiry raised by the interview subjects. Also, by working from

---

[4] For any work, created by an author who died before 1 January 1969, which was unpublished as of

1 August 1989 (when the CDPA 1988 came into force), the duration of copyright will last until 31December 2039. CDPA, Schedule 1, s.12(4).

[5] Vuopala, A., "Assessment of the Orphan works issue and Costs for Rights Clearance" (May 2010), available at: http://ec.europa.eu/information_society/activities/digital_libraries/doc/reports_orphan/anna_report.pdf and Korn, N., (2009) "In from the Cold" JISC & The Collections Trust, available at: http://www.jisc.ac.uk/publications/reports/2009/infromthecold.aspx [Accessed: 1 Sep 2013]

[6] Vuopala, A., (2010) – See page 39, and the example of the National Archives 'Moving Here' project. 45 rightsholders refused permission, but information regarding the reasons for refusal is not available. As to the rightsholders who could not be identified in relation to 385 documents – it is not made clear whether contact details could not be found for these individuals, or if they were but did not respond to permission requests.

[7] A Knowledge Exchange project between The National Archives and University of Glasgow will be able to explore this area in greater depth. More information on the "Technically-assisted sensitivity review of digital public records" project is available at: http://blog.nationalarchives.gov.uk/blog/digital-records-sensitivity-review/ [Accessed: 13 Dec 2013]

a set list of questions, it was still possible to collate the answers to specific questions, thereby producing some quantifiable data. The questions were generated through a discussion of the aims of the project, and through consultation of other similar projects: specifically Jean Dryden's doctoral research on the practices of Canadian repositories when making their archival material available online.[8]

The interviews covered five specific areas: the respondents' role within their place of work; their role working on the Codebreakers project; the rights clearance challenges posed by the project; the development of policies and practices over the timeframe of the project; and finally, the respondents' own views and experience of copyright, divorced from an institutional perspective.

Using this method of enquiry, respondents were asked about policies and practices mandated by the institutions they work for, but also for their own opinions as professionals, built up though work experience, education and training. By encouraging the respondents to reflect on their own experience, it was also possible to use the 'snowball' technique to tease out pointers to documentation and other resources used by the respondents, not only in relation to this project, but to their experience working on others as well.

Once completed, the interviews were transcribed to provide usable data. The transcriptions were used for a variety of purposes. Responses to specific questions were logged to give quantifiable results. Where respondents gave specific examples of practices, or policy changes, these are used as evidence in the text of this report. As a result of interviewing respondents at the Wellcome, project documentation was also made available for inclusion in this case study. This material takes two forms: policy documents which were circulated and updated as the project took place; and an internal report produced towards the end of the project, which includes a variety of statistics and more general lessons learned. The documentation has been discussed in the text of the report and examples of specific documents are available in the Appendix.

This combination of anecdotal evidence, combined with results and project documentation, has allowed us to trace the processes by which the WL has used risk management to achieve their digitisation objectives.

[8] Dryden, J., "Copyright issues in the selection of archival material for internet access" (2008) Archival Science 123-47

2: Introducing Codebreakers: Makers of Modern Genetics

2.1 The Wellcome Digital Library

The WL project commenced in 2010 and since then the Library has developed an online resource for the history of medicine by digitising a substantial proportion of their holdings and making the content freely available on the web. Their aim is to make 30 million pages of archive and library material available online by 2020.

Guided by their transformation strategy,[9] they select material based on the strengths of their holdings and the interests of current or potential audiences. They aim to create significant online resources that will stimulate research in the global health themes that underpin their collecting strategy, and over time plan to digitise key holdings linked to the Wellcome Trust's research challenge areas. These areas cover:

- Maximising the health benefits of genetics and genomics
- Understanding the Brain
- Combating infectious diseases
- Investigating development, ageing and chronic disease
- Connecting environment, nutrition and health

In addition, the WL includes content from other institutions (material that complements the Wellcome's own holdings), and the WL participates in commercial partnerships to achieve cost-effective digitisation of other parts of their collections.

In terms of the type of material being digitised, the WL's digital collections are growing to include:

- cover-to-cover books
- video and audio
- entire archive collections and manuscripts
- paintings, prints, drawings, photographs and ephemera
- born-digital content

As of August 2013, two million images are available via the Wellcome Library Player,[10] a tool specifically developed to display the various types of digital content available on the Wellcome Library website.

Previous digitisation projects undertaken by the WL include: Arabic Manuscripts Online;[11] recipe book manuscripts,[12] Wellcome Film, and AIDS posters.[13] These resources will be linked to the main Wellcome Library catalogue as the digitisation programme progresses.

---

[9] Available at: http://wellcomelibrary.org/about-us/library-strategy-and-policy/transforming-the-wellcome-library/ [Accessed: 11 September 2013]

[10] The Player can be accessed by browsing the Wellcome website or searching their catalogues. For more details, see http://blog.wellcomelibrary.org/2012/11/the-player-a-new-way-of-viewing-digital-collections/ [Accessed: 11 September 2013]

[11] Available at: http://wamcp.bibalex.org [Accessed: 1 September 2013]

2.2 Codebreakers: Makers of Modern Genetics [14]

Codebreakers: Makers of Modern Genetics contains around 1.6 million pages of books and archival material relating to the history of genetics. The project relates to the first of the Wellcome's five key research themes: maximising the health benefits of genomics and genetics. The collections digitised as part of the Codebreakers project include the papers of Francis Crick, James Watson, Maurice Wilkins and Rosalind Franklin – the four individuals most closely associated with the discovery of the 'double helix' structure of DNA in 1953. The archive material includes the papers and records of:

- Francis Crick (1916-2004)
- Fred Sanger (born 1918-2013)
- Gerard Wyatt (born 1925)
- Arthur Ernest Mourant (1904-1994)
- Hans Gruneberg (1907-1982)
- Robert Race (1907-1984) and Ruth Sanger (1918-2001)
- Sir Peter Medawar (1915-1987)
- Carlos Paton Blacker (1895-1975)
- Honor Fell (1900-1986)
- The Medical Research Council Blood Group Unit (1935-1995)
- The Eugenics Society (1863-2008)

The Codebreakers project also includes archive material from external partners that helps place the WL's holdings within a broader context. From the first half of the 20th century the resource includes the papers of J B S Haldane, a leading figure in British science and the first Professor of Genetics at University College London. It also includes the collections of Guido Pontecorvo and his students Malcolm Ferguson-Smith and James Renwick, who helped make Glasgow a leading centre for the study of medical genetics.

In digitising this additional archival material, the WL have worked with five external partners: Cold Spring Harbour Laboratory Library; King's College London; University College London; University of Glasgow; and the Churchill Archives Centre. Each of these external partners digitised archive material from their own collections for inclusion within the WL.

---

[12] Available at: http://archives.wellcomelibrary.org/DServe/dserve.exe?&dsqIni=dserve.ini&dsqApp=Archive&dsqDb=Catalog&dsqCmd=Overview.tcl&dsqSearch=%28%28text%29=%27recipe%20books%20project%27%29AND%28Level=%27item%27%29 [Accessed: 1 September 2013]

[13] Available at: http://www.wellcomecollection.org/explore/sickness--health/topics/aids-posters.aspx [1 September 2013]

[14] The text in this section is an edited version of that found at http://blog.wellcomelibrary.org/2013/03/codebreakers-makers-of-modern-genetics/ and http://wellcomelibrary.org/about-us/projects/digitisation/makers-of-modern-genetics-project/

The books digitisation project consists of over a thousand titles relevant to the Codebreakers project, relating to the science, history and social and cultural aspects of genetics and related disciplines. For the most part, these books were all first published in the 20th century. All of this digitised material, both archives and books, is currently available online. It can be accessed by searching the Wellcome Library catalogue, and digital content can be viewed through the Wellcome Library Player.

In addition to the library catalogue, there are a variety of entry points to the collections through the Codebreakers section of the Library website. Users can browse the digitised collections by subject, discover the background to the individuals and organisations involved, or find out more about the history of modern genetics through specially commissioned essays and reports. There is also an interactive timeline that includes links to selected items from the archives to provide an alternative way in to the subject.

In making these collections available online, the WL engaged in a risk-benefit analysis. The collections are full of personal sensitive data and the copyright in the material belongs to multiple rightsholders, potentially running to thousands of individuals. They have balanced the risks associated with both unknowingly publishing sensitive data, and knowingly publishing copyright-protected material online, against their stated objectives of targeted collection, strategic digitisation and expert interpretation of historical sources, stimulating new research and making the cultural and historical context of medicine accessible to all.[15]

Some archive material has been restricted because of these sensitivity and copyright issues. Access to archive material digitised as part of the Codebreakers project typically requires a log-in to view. However, the WL have tried to make the log-in process as easy as possible for researchers by offering a range of sign-in options, including Twitter, Google, Facebook, OpenID or a Wellcome Library card.

When users log-in, they are asked to accept the WL's Terms and Conditions.[16] From the WL's perspective, these terms and conditions are necessary to ensure that the WL, and by extension the Wellcome Trust, are both complying with data protection legislation and covering some aspects of copyright law. Individual items available on the WL website are subject to specific conditions of use.

The next major archive digitisation project that the Wellcome Library is planning concerns mental health and neuroscience, preparations for which commence in late 2013. Over the course of the next three years the WL will also release other batches of digitised content, including nearly six thousand reports published by Medical Officers of Health in London between the 1840s and 1970s, the complete run of the trade journal *Chemist and Druggist*, 300 medieval manuscripts and up to 10,000 monographs published in the 19th century.

2.3 The Partner Archives

In developing the Codebreakers resource, the WL worked with five external partner archives, each of which digitised material from their own collections for incorporation within the WL. In turn, the partners were: Churchill Archives Centre; Cold Spring Harbour Laboratory Library; University of Glasgow; King's College London and University College London.

Churchill Archives Centre contributed the papers of:

- Rosalind Franklin (1920-1958)

[15] Transforming the Wellcome Library, 2009-2014, available at: http://wellcomelibrary.org/about-us/library-strategy-and-policy/transforming-the-wellcome-library/ [Accessed: 30 Sep 2013]

[16] Available at: http://wellcomelibrary.org/about-this-site/terms-and-conditions/ [Accessed: 1 September 2013]

Since opening in 1972 to store the papers of Sir Winston Churchill, the Churchill Archives Centre (CAC) has amassed the personal papers of more than 570 public figures, including those of Baroness Margaret Thatcher. The rights clearance process required to deliver the Churchill Papers online[17] with Bloomsbury Publishing took almost five years,[18] giving CAC valuable experience of clearing copyright in archive material, which they were able to bring to the Codebreakers project. They have also recently completed the digitisation of the Thatcher Papers.[19] CAC have an international programme of exhibitions, most recently travelling to Ottawa and New York.

Cold Spring Harbour Laboratory Library and Archives contributed the papers of:

- James Watson (1928-)
- Sydney Brenner (1927-)

Cold Spring Harbour Library and Archives (CSHL) is an institutional repository which collects and preserves the "rare books, manuscripts, photographs and scientific reprints documenting genetic research, and the work of the Laboratory faculty, since 1890." James Watson, their president for life, is a strong advocate of open access to, and online publication of research material. Previous digitisation projects include selections of the personal collections of Hermann Muller, Barbara McClintock, Alfred Hershey and Walter Gilbert, an Oral History series and reprints collections, all of which are available on the CSHL Digital Collections site.[20]

Glasgow University Archive Services catalogued and digitised the papers of:

- Guido Pontecorvo (1907-1999)
- James Renwick (1926-1994)
- Malcolm Ferguson-Smith (1931-)

Glasgow University Archive Services (GUAS) collects not only the records of the University, but also houses one of the largest business archives in Europe, charting Scotland's industrial and commercial history. As a result of working with the Wellcome, GUAS have established an internal digitisation unit, and intend to pursue larger digitisation projects in future. Previous projects include digitising the complete run of the Glasgow University Guardian,[21] and both the University of Glasgow Story and the International Story, which commemorate previous graduates and staff of the University.[22] Future projects include the digitisation of Glasgow University Magazine, starting with the period covering WW1, in time for commemoration events in 2014.

King's College London Archives & Information Management digitised the papers of:

- Maurice Wilkins (1916-2004) and the Medical Research Council Biophysics Unit

[17] Available at: http://www.churchillarchive.com/index [Accessed: 29th October 2013]

[18] Adams, N, Packwood, A, & Riley, A. Personal Interview. 9th May 2013.

[19] A selection of the Thatcher Papers is available at http://www.margaretthatcher.org/ [Accessed: 25 May 2013]

[20] CSHL Digital Collections site available at: http://archives.cshl.edu/

[21] Glasgow Guardian Digital Archive available at: http://www.gla.ac.uk/services/archives/guardian/

[22] University of Glasgow Story available at: http://www.universitystory.gla.ac.uk/; International Story available at: http://www.internationalstory.gla.ac.uk/

King's College London Archives and Information Management (KCL) began digitising their collections on an ad-hoc basis in 2003, and have since developed a large collection of digital images. This includes the Serving Soldier resource, funded by JISC, which allows "access to thousands of digital copies of unique diaries, correspondence, scrapbooks, photographs and other archive sources scanned [and] provides an insight into the lives of servicemen and their families from the late nineteenth century until the Second World War."[23] They are also hosting a landmark database project which will provide access to the World War I related holdings of UK university archives, libraries and museums, to coincide with the 2014 centenary.[24] KCL also manages the AIM25 website, a single access point to the catalogues of 120 cultural and scientific institutions in the London area.[25]

University College London Special Collections catalogued and digitised the papers of:

- JBS Haldane (1892-1964)
- Lionel Penrose (1898-1972)
- Francis Galton (1822-1911)

University College London Special Collections (UCL) began their first major partnership digitisation project in 2003. Britain in Print[26] started with the aim of digitising printed material produced in the 300 years following the invention of the printing press. The Joseph Hume Tracts were digitised as a later part of the project, which "reflect the major political, economic and social developments and reforms taking place in Britain in the early part of the nineteenth century, and include some of the causes championed by Joseph Hume during his parliamentary career, such as universal suffrage, Catholic emancipation, a reduction in the power of the Anglican church and an end to imprisonment for debt."[27] Recent projects include Transcribe Bentham,[28] a participatory project which asks volunteers to transcribe digitised pages of Bentham's manuscripts, and encode the resulting text in XML to enable advanced text searching. UCL also continues to digitise and make available items from the George Orwell Archive, the Moses Gaster collection and the papers of Moses Montefiore.[29]

[23] Serving Soldier available at: http://www.kingscollections.org/servingsoldier/home

[24] World War I Discovery database available at: http://www.jiscww1discovery.net/home

[25] AIM 25 available at: http://www.aim25.ac.uk/

[26] Available at http://www.britaininprint.net

[27] Hume Tracts Collection available at: http://www.jisc-collections.ac.uk/Catalogue/Overview/index/992

[28] Available at http://blogs.ucl.ac.uk/transcribe-bentham/

[29] These collections are regularly added to and are accessible from http://digitool-b.lib.ucl.ac.uk:8881/R

3 Archives Digitisation

A brief explanation of the WL digitisation workflow processes is provided; more detailed technical and organisational descriptions are available on the WL website.[30]

3.1 Archives Digitisation at the WL

It is important to note that although Codebreakers was the main project during the pilot programme at the WL, other digitisation projects have been running concurrently, including Medical Officer of Health reports for Greater London and Early European Books (in partnership with ProQuest).[31] The aim of the pilot was to develop the policies, processes and infrastructure necessary for mass digitisation and online delivery to take place, and then integrate those into the existing work processes of the WL.

These range from the complex, i.e. the development of the Digital Player, to the deceptively simple, for example organising the order in which collections are photographed in order to minimise disruption to readers. It will be useful to map out where copyright and sensitivity checking are integrated into the digitisation workflow at the WL. However, in order to do this, some explanation of the digitisation effort, and particularly of the WL metadata processing, will be necessary.

The archival digitisation effort required to bring Codebreakers to fruition presented a number of specific challenges. The first of these concerned the project's scope. Whilst the collections were selected on the basis of provenance and content, very little item-level pre-selection or appraisal of the material was carried out, resulting in the incredible volume of material available on the WL website (2 million images and counting). Whereas large-scale digitisation projects are relatively common in the library sector, the mass digitisation of archival collections (for non-commercial purposes) is still a relatively unusual undertaking.[32] One or two other initiatives might be compared with the Codebreakers project, such as the Welsh Experience of World War I 1914-1918[33] or the BT Archive[34] (both of which were funded under the JISC Mass Digitisation Strand in 2011); however, in terms of the quantity of images being made available online, both of these projects are much smaller in scope and ambition[35] than the WL's programme. The sheer scale of the Codebreakers project demonstrates a large appetite for risk on the part of the WL, in a sector which is known to be highly risk-averse.[36]

---

[30] For descriptions of the ingest process, optical character recognition, using the GOOBI workflow management tool , server capacity and the development of the Wellcome Library Digital Player, see the blog posts available at http://wellcomedigitallibrary.blogspot.co.uk/.

[31] Wellcome's own Genetics collections totalled 600,000 images; the partner archives contributed 500,000 in total; 2000 genetics books have produced 600,000 images; ProQuest are digitising 5.5 million pages of Early European Books on the Wellcome Library site; 7000 London Medical Officer of Health reports have generated 400,000 images; and the second phase of digitisation concerning mental health and neuroscience, will begin in late 2013. The Wellcome Library also accession small amounts of born digital material, but this will continue to grow – Henshaw, C. "Managing Large Scale Digitisation at the Wellcome Library" (presentation, Sync or Sink: Opportunities for Libraries in the Digital Age, Birkbeck College, University of London, 24 Nov 2011) Slides available at: http://www.slideshare.net/Wellcome/cdp25-nov2011 [Accessed: 13 July 2013]

[32] The Churchill Papers (website) provide a good example of the large-scale digitisation of archival material for commercial purposes (made available on a pay-for-subscription basis).

[33] Available at: http://cymruww1.llgc.org.uk/ [Accessed: 12 September 2013] The project includes 190,000 pages of archive material, 50 hours of sound recordings and 20 hours of audio-visual material.

[34] Available at: http://www.digitalarchives.bt.com/web/arena/catalogue [Accessed: 12 September 2013] Over 450,000 pages of archival material were digitised.

[35] Both of the aforementioned projects also involved pre-selection of the material to be digitised; other than selecting specific collections, the Wellcome did not pre-select any of the material included in Codebreakers.

[36] Deazley, R., p25.

The second challenge, therefore, was ensuring that these disparate collections, held between six different institutions, each of which operated with different guidelines and policies in place, would be able to complete the digitisation process, the sensitivity checks and identify relevant rightsholders in the collections.

The third challenge was the fact that many of collections held by the third parties were un-catalogued,[37] or had been catalogued to file level but not to item level.[38] In simple terms, this means that there is either a) no identifying metadata for the digital image, or b) insufficient granularity of metadata for the digital image. Collections have to be catalogued from scratch, expanded, or item level metadata created from the existing catalogues.

In addition to the cataloguing required, the collections had to be processed to the point where they were able to be digitised: the cleaning, removal of bindings, unfolding, repackaging and marking up of material that constitutes the most basic of preservation treatments. This process allows the material to photographed or scanned at speed, and a unique identifier attached to the resulting digital images.

The fourth challenge concerned taking material out of circulation for extended periods of time, while conservation, labelling, cataloguing and digitisation work could be carried out. Letting readers know that material will be unavailable for consultation in advance of such an occurrence ensures smooth delivery of services and that visitors can plan research trips to the archive accordingly.

3.2 The WL Digitisation Workflow[39]

At the time of writing[40], the WL had created 3.9 million digital images (two million of which were available online), with 6-7 on site full time photography contractors and 8 full and part-time support staff. During the pilot, there were 3 Content and Metadata Officers (CMOs) who specifically worked on Preparation and Retrievals, Sensitivity Checks, and Copyright Clearance. In order to ingest the images and metadata generated at the 3rd party partner archives, there was also an archivist assigned to the project team to provide metadata support.

Once the images are created, they go through a series of processes, some of which can be seen in Figure 1. There are 2 Digital Ingest Officers who control the processing and ingest of the digital images using Goobi, a workflow management tool. Goobi automatically generates METS files which provide administrative and structural information to the Player to display digital content, and which can be manually edited to give instructions for the licence and access levels.[41]

The metadata in the METS is used to provide content description and create the structure which users can see on the Player – covers, tables of contents, title pages and pagination. Once the images and associated metadata have been edited in Goobi, the images are stored in the WL's permanent storage, Safety Deposit Box, which facilitates long-term preservation, future migration, and acts as the gateway to securely-stored content. Given the decision by the WL to digitise everything contained in the selected collections, it was considered impractical to remove single documents because of refused permission or sensitive data. As a result, those images will

---

[37] Furlong, G, Makin, K, & Wright, S. Personal Interview. 6th June 2013.

[38] Browell, G. & Methven, P. Personal Interview. 28th May 2013.

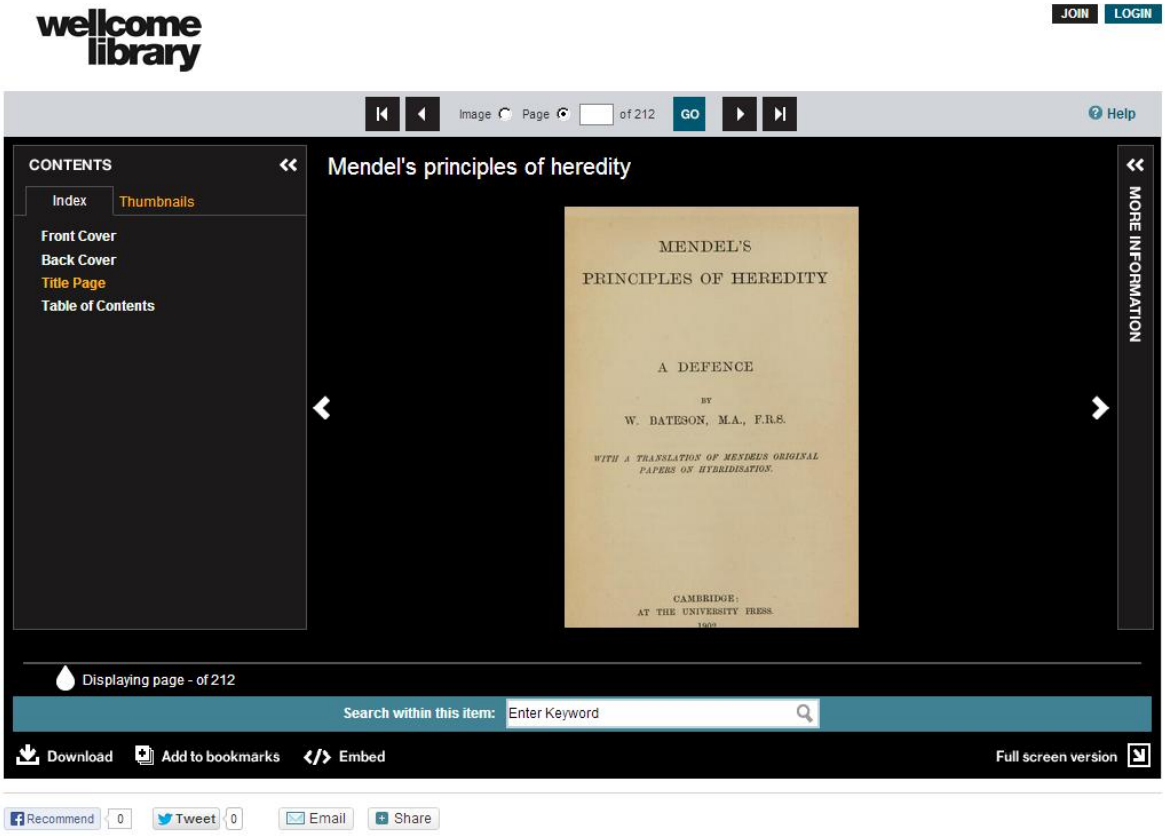
[39] Christy Henshaw is the Digitisation Programme Manager at the Wellcome Library. For a full, technical description of how the Wellcome Library has seamlessly integrated digitisation into their online presence, see Henshaw, C. & Kiley, R., "The Wellcome Library, Digital". July 2013, Ariadne Issue 71. Available at: http://www.ariadne.ac.uk/issue71/henshaw-kiley [Accessed: 20 July 2013].

[40] This report was drafted between August and November 2013.

[41] Licences relate to the access options offered to rightsholders of library material - this is discussed in more detail in Section 4.1.1. Access levels are determined by the sensitivity of the information contained in the archive material, and whether permission has been given to publish the archive or library material in question. Access levels are discussed in more detail in Section 5.2.

be stored until they are either out of copyright or no longer deemed to be sensitive. Once the images have been made available to the Player (Fig 1), links are added to the catalogue records, making the images discoverable.

Figure 1: The Wellcome Digital Library Player



To start the copyright checking process, names of rightsholders were extracted from the catalogue records for each collection in early 2011[42]. This process started independently of digitisation (which began in 2010), and continued to run in parallel for most of the project. The details of the rightsholders were recorded in an Access database, and the results of the permission letters were also recorded there. If permission was given or refused, this information was recorded in the metadata, which allowed the Digital Ingest Officers to amend the METS file, either giving or denying access to the images.

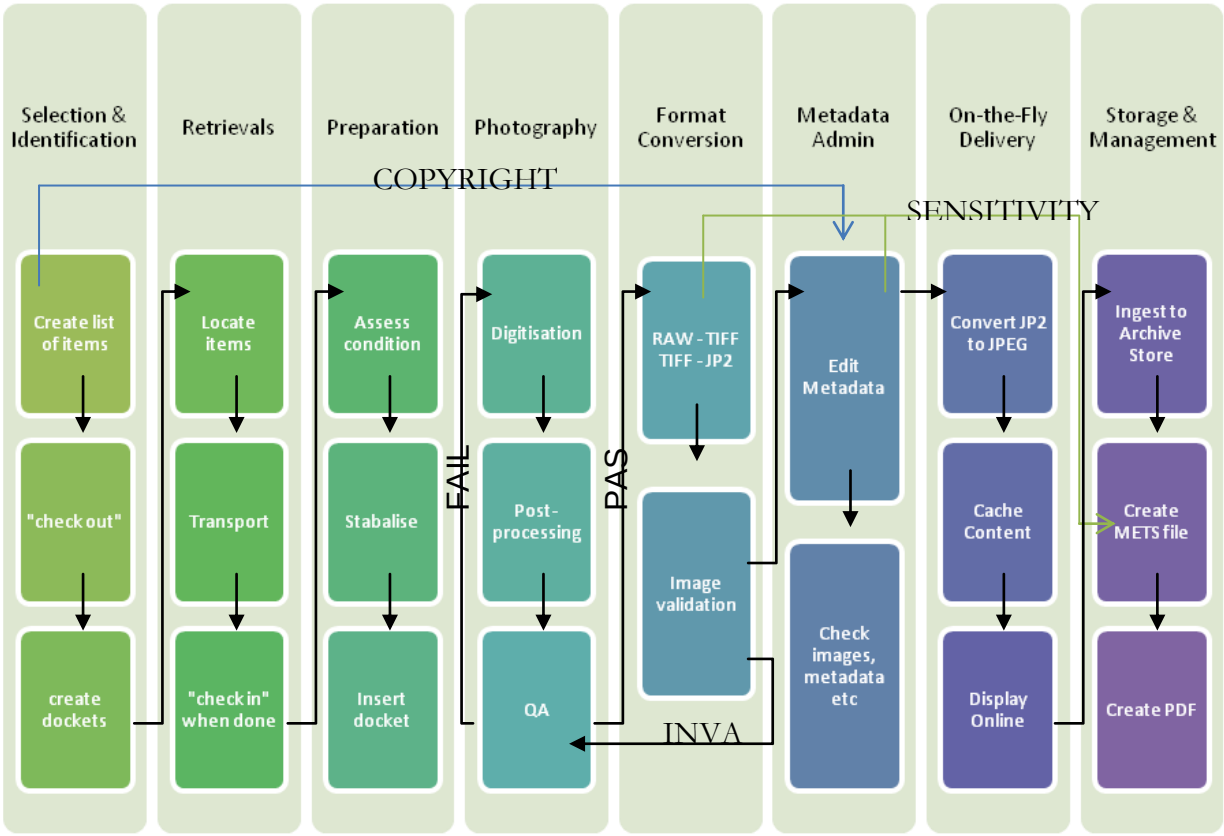
Later in the project, once permission letters had been sent out, the library began to receive enquiries from the rightsholders, who wanted more information on the material that the library held. At this point, a selection of the material was sent to the rightsholder, but this was only possible if the collection had already been digitised. A simplified diagram of the workflow process shows how each of these elements combine. (Figure 2)[43].

---

[42] Relying on catalogue records to identify rightsholders without checking the original documents can potentially cause problems, depending on how the collection was originally catalogued. For example, it is not always standard (or possible, given time constraints) to index every personal name which appears in a collection (if you accept that names appearing in collections are likely to include rightsholders). Also, if the catalogue description is not sufficiently clear, or if read incorrectly, it can be assumed that documents or correspondence which mention a specific person were actually authored by them. One or two examples of this occurred during Codebreakers.

[43] This diagram was originally used in a presentation by Henshaw, C. "Workflow Thinking for Digitisation," (presentation, Digitisation Doctor, Wellcome Trust, London, 15 April 2013). Slides available at: http://prezi.com/unqxvaz5swvp/workflow-thinking-for-digitisation/ [Accessed: 1 July 2013]

Figure 2: WL Digitisation Workflow



4 Copyright

In this section (and in the sections that follow) we draw a distinction between copyright compliance issues raised by library material (published books) and the archival material that made up the bulk of the Codebreakers initiative (the personal papers and records of Crick, Watson, Wilkins, Franklin, and so on).

4.1 Copyright in Library Material

As part of the Codebreakers project the WL decided to digitise up to 2025 books covering various aspects of the study of genetics. The WL intended to clear rights, where possible, in all of the books selected for digitisation, and did not initially employ risk level criteria such as those drawn up for the archive material to reduce the number of rightsholders, as it was anticipated that *all* identified rightsholders would be contacted. It is also important to note that, in order to make the digitisation process as productive as possible, all of the out-of-print titles were digitised regardless of status, although status and permissions determine whether that digital copy is available on the WL website. In some cases, in-commerce titles were digitised before their status was known; these were deleted entirely from the system.[44]

In order to make a digital copy available online, they had to identify whether the book in question was:

- In/out of copyright
- In commerce/out of print
- owners identifiable/orphaned works (for in-copyright, out-of-print titles)

Once the status of a work was identified, an attempt would be made to gain permission to publish with the relevant rightsholder (whether a publisher and/or the author or his or her estate).

4.1.1 Library Clearance Results

As previously stated, the WL identified 2025 published books on genetics which they intended to digitise and make available online through Codebreakers. In order to clear rights in these works, the WL worked with two collecting societies: the Publishers Licensing Society (PLS) and Authors' Licensing and Collecting Society (ALCS). Previously, the Wellcome Library had worked with ALCS, PLS and the British Library on the ARROW initiative (Accessible Registries of Rights Information and Orphan Works), a "network of databases and rights registries designed to enable the identification and rights clearance of works to support mass digitisation throughout Europe."[45]

It was decided that books found to be in-commerce (that is: still in print, and available for sale) would not be published on the WL website,[46] as publication would constitute clear infringement and the availability of digital copies might have a negative impact on the market for such works.

---

[44] Books where permission to digitise was expressly refused were removed from the retrieval list prior to digitisation where possible, but in most cases had already been retrieved and digitised before replies to permission requests were returned. Caroline Herbert, correspondence with the author, December 2013.

[45] Available at: http://pressandpolicy.bl.uk/Press-Releases/Electronic-clearance-of-Orphan-Works-significantly-accelerates-mass-digitisation-524.aspx [Accessed: 30 Sep 2013]

[46] In-commerce books were removed from the digitisation workflow where possible, but it is likely that some have been digitised. They will not be available on the WL website.

Alongside ARROW, the Bowker Books in Print website was used to check whether books were in-commerce in non-ARROW countries, resulting in a total of 252 works being identified as in-commerce.

The list of remaining books was then sent to ALCS and PLS to run through ARROW. Using ARROW revealed that catalogue records do not always contain all of the rightsholders associated with a work, particularly for scientific publications which often have multiple authors. Given the quantity of potential rightsholders this discovery revealed, it was decided that the overall number of rightsholders would have to be controlled in some way. In response, ALCS developed a 3/3 rule that saw authors ruled out of the rights clearance process if they appeared less than three times, in titles with three or more authors. This shows that, even in situations where prior agreements state that rights clearance will be comprehensive, some measure of risk management had to be employed to keep the number of rightsholders to a reasonable quantity.

ALCS and PLS would then inform the WL of the status of each book on the long list, i.e. whether they were: in or out of copyright, an orphan work, or had traceable rightsholders. ACLS and PLS then contacted the traceable rightsholders to enquire about including the work within the Codebreakers project. At this stage, permissions and refusals were returned directly to the WL, so that, where possible, works could be made available online. The WL, ALCS and PLS would share data intermittently, in order to identify and follow-up on rightsholders who did not respond to permission requests.

As part of the permission request, the WL offered rightsholders a number of different options regarding the availability and subsequent use of the work in question. That is, rightsholders could opt to limit the user from (i) downloading the whole work; (ii) sharing copies of the work in whole or in part; or, (iii) altering, adapting or translating the work.[47]

Those works that were designated as orphan works were digitised and made available online subject to a takedown policy (see section 5.1.1 for further details). Where rightsholders were traced, but permission requests did not generate a response, the Wellcome have decided to make these works available online in stages, based on age, with downloads prohibited and again, subject to the takedown policy.

[47] Examples available in the Appendix.

The results of the books clearance process can be seen in Table 1:[48]

Table 1: Results of the WL Library Copyright Clearance Process

Status	No. of Works	% of Total
Out of Copyright (In commerce or out of print)	297	14
In Copyright, In Commerce	252	13
In Copyright, Out of Print		
Permission granted	480	24
Permission denied	206	10
Outstanding Queries	210	10
Did not Respond	375	19
Orphan works	205	10
Total of Books WL cannot make available	**458**	**23**
Total of Books WL can make available	**982**	**48**
Total of Books WL could make available, under a risk-managed approach	**375**	**19**
Total	**2025**	**100**

Comparing the results of the Codebreakers project with other similar projects, such as the British Library (BL) rights clearance study of 140 books published between 1870 and 2010,[49] is illuminating, given that both used the ARROW system to manage the rights clearance process, and especially when the scale of Codebreakers is taken into account, at almost ten times the number of works as the BL study.

Of the titles in the WL sample, 14% were found to be in the public domain, compared to 29% in the BL project. The difference in these figures might be explained by the fact that the majority of the books the WL selected for digitisation were published in the 20th century, and therefore are less likely to be out of copyright.

252 books in total (13%) were eliminated from the project because they were found to be in-commerce. 63%, or 1266 works were in copyright but out of print, a figure which differs from the 79% found in the BL sample, but this can be explained by the outstanding queries, all of which were identified by ARROW as being in-copyright but out of print. If the figures are adjusted to include these titles, the percentage would rise from 63% to 73%[50].

Permission was granted to the WL for 24% of the books, compared to just 17% of those at the BL. One possible explanation for this slightly higher permission rate at the Wellcome is the

[48] Correct as of 27th February 2013 - Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].

[49] Stratton, B., (2011) "Seeking New Landscapes: A rights clearance study in the context of mass digitisation of 140 books published between 1870 and 2010," British Library, London, UK. Available at: http://pressandpolicy.bl.uk/imagelibrary/downloadMedia.ashx?MediaDetailsID=1197 [Accessed: 1 June 2013]

[50] This is accepting that the titles for which Elsevier have refused permission are neither public domain works, or in-copyright, in-commerce titles. The use of ARROW would have identified otherwise.

concentrated nature of the collection: by focusing only on genetics, they were primarily contacting scientists and academics in the field, a group that may be more positively disposed towards the aim of open access to research than, for example, commercial authors. By contrast, the BL project used a randomly generated sample of works[51]. The BL was further constrained by the time limit they had set for the project.[52]

Permission to publish was refused for 206 works, 10% of the overall total. Some of the reasons given for refusals included: publishers were about to digitise the title; a new edition of the title was due for release; or authors simply did not want to be included.[53] Permission was refused, or could not be agreed, for 27 works in the BL study, or 19% of the total.

Permission was also refused for a further 210 books by Elsevier. However, there is some uncertainty as to whether Elsevier actually hold the rights to these titles; the WL is currently working with Elsevier to resolve these outstanding queries. Should the refusal of permission by Elsevier stand[54], the rate of permissions denied would rise to 21%, or 416 works in total. This figure corresponds more closely to the results of the BL study, which found that associations and voluntary group were the types of publisher most likely to grant permission, with museums, schools and private institutions least likely. The BL study also found that commercial publishers granted permission for 6 (15%) out of 39 in-copyright works, but refused permission and did not respond to permission requests for 27 (69%) out of the same 39 works. It is likely that the works chosen by the WL for digitization would have included various types of publisher, but this information was not recorded systematically for review purposes.

Non-response to permission requests were 19%[55] and 26% at the WL and BL respectively. Other rights clearance projects, for both books and archives, have encountered similar problems with non-response. For example, Carnegie-Mellon University Libraries used a random sample of books from their collection to test the feasibility of obtaining permission to digitise and publish in-copyright works online. Of the publishers they identified and contacted, 36% did not respond to repeated permission requests.[56] The Wellcome Library's own AIDS posters project only managed to obtain permission from 16.5% of the rightsholders identified in the collection.[57]

When faced with a situation of non-response from a known rightsholder, after 2 further follow-up attempts, the WL decided to make the material available online subject to their policy of take-down on request. This is a highly unusual policy for a library or an archival institution to adopt. Archivists will typically withhold material from a digitization project if the rightsholder does not respond to a permission request.[58] Consider, for example, the digitization of the Jon Cohen

---

[51] Stratton, B., (2011) p.6

[52] The BL project was limited to 11 months in total, covering the random sampling and selection of 140 titles, determining the copyright status of those works, identifying the rightsholders, attempting to contact the rightsholders, and recording and reporting the results of this process. It is worth noting that, like the WL, the BL study made initial contact (where details were found) to gain permission, and then made two further follow-up attempts if required. Any permissions received or negotiations concluded after the BL project term were not included in the results.

[53] Herbert, C., Personal Interview. 16th April 2013.

[54] There are two possibilities linked to this refusal – Elsevier are the legitimate rightsholder, and do not grant permission, or they do not hold the rights to the works in question, and therefore cannot grant permission.

[55] This is the same proportion of non-response as for the archive permission requests.

[56] Covey, D.T., (2005) 'Response to Notice of Enquiry about Orphan Works,' Carnegie Mellon University Libraries. Available at: http://www.copyright.gov/orphan/comments/OW0537-CarnegieMellon.pdf [Accessed: 15 August 2013]

[57] Vuopala, A., "Assessment of the Orphan works issue and Costs for Rights Clearance" (May 2010), available at: http://ec.europa.eu/information_society/activities/digital_libraries/doc/r eports_orphan/anna_report.pdf

[58] Deazley, R., and Stobo, V. (2013) Archives and Copyright: Risk and Reform. Working Paper. CREATe / University of Glasgow, Glasgow, UK.

AIDS research collection at the University of Michigan. 13,381 items in the Cohen Research Collection were selected for digitisation, but 1973 items of copyright-protected material (36%) were not made available online. It was decided that 981 of the copyright-protected items (18%) would not be displayed, as rightsholders had not responded to permission requests. Compared with 687 items (13%) that were deemed to be orphan works, and 294 items (5%) for which permission was refused, the largest proportion of withheld items were those where rightsholders did not respond to permission requests.

The WL requests for permission to publish online which did not illicit a response will be made available in three batches – the first, which included titles published up to 1930; the second, from 1930 until 1970; and a third batch, from 1970 up to the present day, which caused the greatest concern to the Wellcome Trust's legal team[59]. The most recent batch is regarded as most problematic because the rightsholders will almost certainly still be living, and therefore could be more likely to object to publication without permission.[60]

The total number of orphan works held in library collections has been estimated to be between 5-10%.[61] This was born out by Codebreakers where 10% of relevant works were found to be orphaned. By comparison, 31% of the in-copyright material in the BL study was found to be orphaned. This is perhaps the most surprising divergence between the two projects. It cannot be explained by the BL project's inclusion of works from the late 19th century (on the basis that the older a work is, the more likely it is to be an orphan), as the BL found that the highest proportion of in-copyright orphan works included in the study were works published in the 1980s. The higher proportion of orphan works in the BL study may, however, be attributable to the inclusion of self-published works in their sample (if we accept that a self-published work is more likely to become orphaned).[62] The likelihood of the WL sample including self-published scientific works would be reasonably low.

In total, of the 2025 works identified by the WL for inclusion in the project, 1357 have been digitized and will be made available online (including works that are out of copyright, orphan works and titles for which permission was granted).

Both projects used the ARROW system to identify rightsholders. During the BL study, it was shown that the ARROW system returned 92%[63] of the same results as a manual diligent search, but at a rate of 5 minutes per title, rather than 4 hours.

Estimating the time taken for the WL to clear rights in both books and archives is problematic – Codebreakers, as a project, was not treated as an exercise in which such things would be recorded and evaluated, unlike the BL study, which was designed specifically for that purpose. The CMO responsible for copyright clearance has explained that the time taken to trace varies greatly between rightsholders, and that during Codebreakers, there were other projects running simultaneously in which she had a role.[64] The nature of the project also meant that tracing rightsholders would take place in focused bursts of activity, rather than running in the

---

[59] Bird, C. Personal Interview. 18th April 2013.

[60] Bird, C. Personal Interview. 18th April 2013.

[61] Korn, N., (2009) "In from the Cold" JISC & The Collections Trust, available at: http://www.jisc.ac.uk/publications/reports/2009/infromthecold.aspx [Accessed: 1 Sep 2013]

[62] Stratton, B., (2011) p.41

[63] Stratton, B., (2011) p.5 After system improvements and a manual comparison exercise, accuracy of the ARROW system compared to manual search was improved from 51% to 92%.

[64] Henshaw, C. in Deazley, R., and Stobo, V., (eds) (2014) *Archives and Copyright: Developing an Agenda for Reform*, Digital proceedings of RCUK symposium [Forthcoming]

background as a continual process.[65] This makes calculating average times for rights clearance virtually impossible.

The emphasis the WL placed on the non-commercial nature of Codebreakers also appears to have worked very successfully, given the fact that no rightsholders in the library material asked for payment in return for permission to publish. One rightsholder asked for a donation to a charitable trust, which the WL agreed to honor. In the BL study, which was also conducted on a not-for-profit basis, one rightsholder requested fees. Given that the BL study finished before the percentage of permissions received could rise above 17%, it is difficult to say whether more rightsholders would have requested fees or not.

In a similar random sample rights clearance study by Carnegie-Mellon University Libraries (CMUL), which was also conducted on a non-commercial basis, 6% of the publishers contacted who granted permission requested a fee, ranging from $50-$300.[66] This equates to 4 titles out of 66 for which permission was granted. The study also found that publishers who requested fees were also more likely to place restrictions on the length of the license granted. Both the BL and CMUL studies emphasized their public interest missions and non-commercial intent (with CMUL going so far as to provide evidence in each permission request letter from the National Academies Press that open access to its books did not decrease sales), yet still managed to pick up (admittedly very few) fee requests, while the WL have avoided any such request.

4.1.2 Library material conclusion

The WL have stated that they will not attempt to clear rights in such a large selection of in-copyright books again, as they felt the process was too complex and resource intensive for the return generated.[67] In some respects this is surprising, given that the WL process returned marginally better results than other comparable projects, particularly in relation to permissions received, and the low overall number of orphan works.

It is important to draw one more distinction between the WL and BL/CMUL rights clearance processes - the BL and CMUL were working on the understanding that digitisation would not begin until express permission to digitise had been received, and this was one of the reasons set forth for limiting the time spent on the BL study to 11 months: a definite cut-off date would allow the digitisation process to begin. The WL approach was completely different: given the scale of the undertaking, the WL decided to begin digitisation straight away, regardless of which permissions, if any, had been received. If rightsholders refused permission, their works were either removed from the workflow prior to digitisation, or the resulting digital images were deleted from the WL systems. The BL/CMUL approach is compliance-driven and therefore far more rigid, whereas the WL approach is more pragmatic, risk-assessed and therefore flexible. Technically, in digitising library material before permissions were received, the WL has infringed copyright; however because copies of rightsholders material have not been made publicly available, the risk of objection is considered to be low.

4.2 Copyright in Archive Material

The material in the archive collections selected for digitisation varied greatly, from textual materials including personal and institutional correspondence, scientific notes, writings in draft, laboratory notebooks, teaching files, papers relating to scientific meetings attended, biographical

---

[65] Henshaw, C. (2013) Archives and Copyright: Developing an Agenda for Reform, Q&A Session.

[66] This figure is taken from Page 14 of the Digital Library Federation Report by Covey, D.T., (2005) *Acquiring copyright permission to digitise and provide open access to books,* DLF, Council on Library and Information Resources, Washington DC. Available at http://www.clir.org/pubs/reports/pub134/reports/pub134/pub134col.pdf [Accessed: 20 Sep 2013].

[67] Henshaw, C., Personal Interview. 16th April 2013.

material, diaries, lecture notes, research papers, article reprints and newspaper clippings, to administrative material including project documentation, records, and minutes, to audio and visual material including photographs, posters, slides, publicity material, interviews, graphs, data sets, x ray diffraction photographs and audio recordings.

Given that all of the archival material selected for digitisation was 20th century in origin, and that the vast majority of it was unpublished,[68] at no point did the WL consider checking for material in which copyright had expired. That is, the archival collections were treated as being in-copyright.[69]

The WL separated the rightsholders in the collections into two categories:

- Creator copyright holders (in this report, 'creator copyright' is defined as copyright in material created by the subject of the collection - the original creator or organisation, with rights inherited by heirs of the original creator or succeeding organisations, or signed over to the archive on deposit)
- 3rd party copyright holders (where material in the collections had different or multiple authors)

The WL decided to clear rights with the creator or creating organisation of each collection, or with their successors, estates or heirs, where necessary. In some cases, where the creator was still living, as in the case of Watson and Brenner, this was a simple formality. For other collections, like those of Renwick and Haldane, who were both dead, it was more problematic as heirs had to be traced. Clearing creator copyright was not a risk-managed process; rights had to be cleared.

All of the collections selected contained large amounts of correspondence, with potentially thousands of third party rightsholders overall. Indeed, Wellcome have estimated that about 3.9 million images of archive material were created as part of Codebreakers, a high proportion of which will have involved third party copyright owners. Instead of contacting every potential rightsholder, the Wellcome opted only to contact those they felt posed a risk of objecting to the publication of their material. Wellcome also decided, on the basis that the material had been created in a non-commercial capacity, that no licence fees would be offered to rightsholders who were contacted and asked for permission to publish. This policy was adopted from the start of the project.

Later in the project, the WL also decided to block online access to some published material present in the archive collections[70] – for example, large proportions or entire issues of magazines, journals and grey literature. This decision was a combination of both pragmatism and risk management. In many cases, copies of the material were available elsewhere, in digital form, and the WL provides links to this material where possible. Including published material would have increased the number of rightsholders whose permission would be required for online access.

---

[68] For any work, created by an author who died before 1 January 1969, which was unpublished as of 1 August 1989 (when the CDPA 1988 came into force) the duration of copyright will last until 31 December 2039. CDPA, Schedule 1, s.12(4).

[69] Herbert, C., Personal Interview. 16th April 2013.

[70] This decision appears to have been taken in response to staff confusion about which published materials it was acceptable to make available online - Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].

4.2.1 The Collaboration

At the beginning of the project, the 3rd party archives were invited to two roundtable meetings, the first in September 2010 and the second in January 2011.

At the first meeting, when considering rights clearance procedures, two possible scenarios were discussed: in the first, partners would be responsible for contacting the rightsholders of material held in their own collections; and in the second, this process would be administered centrally by the WL. After discussion, two decisions were taken. First, the WL would have central responsibility for contacting 3rd party rightsholders, as this would avoid potential problems such as one rightsholder being contacted by multiple institutions; and second, the partner archives would clear rights with the original creators (or their families) for the collections they held.

The WL's risk criteria for determining which rightholders should be contacted were also discussed at this meeting,[71] as a document containing the original risk criteria had been circulated prior to the meeting. These risk criteria were modified slightly in response to discussion with the partner archives. For example, KCL and GUAS felt that the risk criteria had been written with library material in mind, and that the criteria was focused on the possibility of legal action being taken against the Trust, rather than the risk of reputational damage, or maintaining good relationships with depositors.[72]

Churchill Archives were also keen that their participation in Codebreakers should be consistent with their own previous copyright clearance strategy on the Churchill Papers. Early discussion with the WL, before Churchill had agreed to join Codebreakers, had left Churchill with the impression that only the creator's copyright would be cleared in the collections, and they were keen that significant rightsholders present in the collections should be contacted for permission as well. They were not expecting the WL to engage in the same level of comprehensive rights clearance which they had to negotiate for the Churchill Papers[73], but consistency of approach was an important consideration for them during the project negotiations.

This emphasis on maintaining good working relationships with rightsholders and depositors was reflected by the WL themselves, who contacted former Wellcome Trust board members where their names were listed in the metadata.[74] However, these adjustments to the risk criteria were not explicitly added to the documents, and it was left to each partner archive to decide whether they wanted to contact these other stakeholders themselves, or if they wanted that contact to be made through the WL.

With the risk criteria now agreed, the 3rd parties were able to compile a longlist of rightsholders identified in each collection, and several of them had completed this exercise by the time of the next roundtable meeting, in January 2011. The longlisting exercise will be examined in more detail in the next section.

4.2.2 Archive material clearance results

As we have discussed, the WL held central responsibility for contacting 3rd party rightsholders, and also defined a set of risk criteria that they and the partner archives would use to identify high-risk rightsholders to be contacted as part of their risk-managed approach to copyright compliance. This section considers the WL's copyright risk management strategy in more detail.

In 2008, the Wellcome Library digitised a collection of 3000 AIDS related posters, which they made available through Wellcome Images. This project involved their most concentrated

[71] See page 27 and the Appendix for more details.

[72] Browell, G. & Methven, P, Personal Interview. 28th May 2013.

[73] Clearance for the Churchill Papers was extremely comprehensive and took approximately 5 years.

[74] Former Wellcome Trust board members were contacted if there was any mention of them at all in the collections - the Wellcome Library applied a lower threshold of risk because the Board were known to them and they wished to manage their internal reputation/ courtesy. Sue Davies, correspondence with the authors, December 2013.

copyright clearance effort to date, and the best endeavours approach to searching for rightsholders developed during this project influenced the way in which copyright compliance was managed for Codebreakers.

For the AIDS posters project, the Wellcome Library took the view that the posters were originally created to raise awareness of AIDS-related issues, and that few organisations would object to the images being posted online.[75] The rights management workflow was fairly straightforward. Each poster was checked for identifying marks. If no ID was found, the image was put online. If a creator or organisation could be identified, a search for contact details was conducted. If the search was fruitless, the material was posted online. If a creator was contacted and refused permission, the image was not included in the online resource. If after two months, the creator had not responded to the request, the image was made available online.[76]

It was during these early digitisation projects that the Wellcome Library also first developed their takedown policy,[77] a policy designed to mitigate some of the risks associated with publishing material online, which can either be in-copyright, or potentially contain sensitive or defamatory information. If a rightsholder complains about material on the site it is taken down immediately, while the complaint is dealt with; this allows the institution to manage the situation as it develops, and demonstrate it has acted quickly and in good faith.[78]

Whilst the search policy for AIDS posters was relatively straightforward (an attempt was made to identify every copyright holder represented in the collection), the WL understood that this approach would not be workable for Codebreakers, given the scale of the project. Instead, they opted to specify a set of criteria that could be used to reduce the overall number rightsholders to be included in the rights clearance process.

The risk criteria developed for the Codebreakers project consisted of three levels: Low, Medium and High. Low Risk was a default category into which all material not deemed to be Medium or High Risk fell. The categories of Medium and High Risk were defined as follows:

Medium Risk: When all of these three factors apply the material is classed as medium risk and should be added to the long list,

- The author/creator has (or had) a high public profile.
- The author/creator is alive or is known to have a literary estate (as recorded in http://tyler.hrc.utexas.edu//)
- The material appears to have been published/broadcast and/or prepared for commercial gain, rather than to advance academic knowledge or in a not-for-profit environment

High Risk: If **any one** of the following apply the material is classed as high risk and should be added to the long list,

---

[75] Henshaw, C., Personal Interview. 18th April 2013.

[76] Wellcome Library, (Feb 2008) AIDS Campaign Posters – Copyright management strategy [Accessed: 23 April 2013]

[77] The Wellcome Takedown Policy is available at: http://wellcomelibrary.org/about-this-site/copyright-clearance-and-takedown [Accessed: 01 Aug 2013]

[78] For more information on the use of takedown policies, see Hubbard, B., (May 2008) "Take Down Policies" JISC Repositories Support Project, available at: http://www.sherpa.ac.uk/documents/take-down.pdf [Accessed: 2 September 2013]

- The author/creator is a well-known literary figure, broadcaster or artist.
- The author/creator/literary estate/publisher is known to actively defend their copyright
- The relationship between the holding institution and the author/creator/publisher is awkward
- There is a large proportion of material from an author/creator which appears to be in copyright, i.e. more than 20% in a box[79]

A number of comments can be made about these various criteria.

In terms of Medium Risk, judging the first criterion requires specific subject knowledge. Without such knowledge, archivists rely on sources like Who's Who and Wikipedia, neither of which offers a reliable indication of how 'well-known' a particular person is, other than the fact of their inclusion in such a resource. The second criterion specifies the use of the WATCH file, an excellent resource; however it is not comprehensive, and again, it is based on published works and their authors. The third criterion is a useful measure, but archives are unlikely to contain high proportions of published material, and knowledge of the subject's career, again, is likely to be more appropriate.

For High Risk, the same points for the first criterion apply. For the second, which sources should an archivist consult to gauge whether someone actively defends their copyright? News resources may be useful, or a name-check of case law, but would such disputes come to trial? The third criterion plays directly to the concerns of the partner archives concerned to maintain good relationships with their depositors; such difficulties are not uncommon and should be considered carefully.

There are two points to make about the fourth criterion, particularly when we bear in mind that only one criterion is enough for inclusion on the long list. First, it states "material from a creator which **appears** [emphasis added] to be in copyright," when WL practice[80] was to treat all the collections as if the material they contained was still in copyright, given their age. Second, archivists at GUAS felt that proportion of material was an illogical factor given that it could take only one document by someone especially well-known (for example, Winston Churchill) to provoke a complaint/takedown notice/suit.[81] That is, if someone is in the habit of defending their copyright, it doesn't matter how much of their material you've got: they will complain regardless.[82]

As previously discussed, the WL and the 3rd parties had agreed on risk criteria at a roundtable meeting in September 2010. By the time the project partners met again in January 2011, the 3rd parties had compiled longlists of rightsholders identified in their collections, lists which they had composed using the WL risk criteria. The WL was then able to cross check all of these lists to produce a final longlist – a list which was considered to be too lengthy.

This list, through an iterative, negotiated process between the WL and the 3rd parties, was then reduced by nearly two-thirds, from approximately 450 names to 160. This suggests that the 3rd

---

[79] Wellcome Library, (Nov 2011) Guidance on Assessing Copyright Risk for Archives, Internal Document [Accessed 20 April 2013] Available in the Appendix.

[80] Herbert, C., Personal Interview. 16th April 2013.

[81] Richmond, L., Personal Interview. 26th April 2013.

[82] Whilst the fourth criterion appears in internal documents, in practice, the Wellcome Library team did not use it as a deciding factor during the long-listing and short-listing process for identifying rightsholders in the archive collections.

parties struggled to apply the risk criteria in the way that the WL initially anticipated. Possible reasons for the length of the lists could be the breadth of the criteria in relation to the specialised nature of the collections, or that the 3rd parties may have been slightly more risk averse than the WL expected and didn't apply the criteria with the expected ruthlessness.

With the final list of rightsholders now in place, the search could begin. The WL team used a variety of resources to track down rightsholders:

- Who's Who
- The WATCH File
- Google
- Wellcome Trust internal databases
- Third party archives
- Dictionary of National Biography, obituaries, Wikipedia, to identify descendants and heirs (finding addresses required using sources above)[83]

The CMO with responsibility for copyright clearance noted that some contact details "only surfaced because the clearance effort for the library material was taking place simultaneously."[84] The WL had no hard and fast rule for how long to spend searching for a rightsholder: this was left to the discretion of the Content and Metadata Officer and the External Projects Officer.

If a rightsholder's address or the address of a partner or descendant could be found, a letter would be sent requesting permission to publish their material. The letter contained the following:

- A statement explaining the WL mission and its role within the Trust;
- a paragraph outlining the Codebreakers project and what will be published online;
- a paragraph summarising the material held in the collection, the copyright in which may belong to the addressee, and a request to digitise it;
- a statement explaining the potential of the collections for research and outlining the Wellcome policy on sensitive personal data, with a link to the Access to Archives Policy;
- a Copyright Permission Form;
- a Copyright Refusal Form, and
- A list of Frequently Asked Questions.

---

[83] Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].

[84] *Ibid*

Examples of these can be found in the Appendix.

The copyright permission form is a straightforward request to use the material in the project. Unlike the way in which the WL negotiated permissions in relation to the use of library materials, copyright owners in archival material were not presented with options to block download of whole or parts of the works, and so on. It was decided by the WL that managing such permissions in relation to archival material would have been too administratively burdensome.[85]

As the project went on, it became clear that including printed samples of the material in question,[86] rather than relying on a short description of the work (which was likely to spark a series of enquiries), was an effective way of communicating with and securing permissions from rightsholders, although time-consuming to achieve.[87]

Letters were sent out in batches, as and when details were found, and if no permission or refusal forms had been returned after a delay of two months, follow up letters (and emails, where possible) were sent. If rightsholders did not respond to a second permission request, the work in question was re-assessed for inclusion in the project according to three new risk levels[88]:

- Low Risk and known to the WL[89]: put the material online
- Low Risk: put online after two month delay (generally scientists and politicians)
- High Risk: not to go online without express permission (commercial authors, dead or alive)

The results of each stage of the rights clearance process were managed and recorded using an Access database.

The results of this process can be seen in Table 2:[90]

[85] Henshaw, C. Personal Interview. 18th April 2013. On two occasions, the Wellcome was forced to accept a five-ten year license from rightsholders – in practice this was discouraged by emphasising the option in the takedown policy of requesting the removal of the material at any time.

[86] Observe that those printed samples are technically infringing copies.

[87] Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].

[88] These 'extra' levels of risk were defined by Richard Aspin, Head of Research and Scholarship at the Wellcome Library, in order to deal with the number of non-respondents in the build up to the launch of the Codebreakers site.

[89] Christy Henshaw has suggested that 'known to the Wellcome Library' would be more accurately described as 'directly associated with the Wellcome Trust.'

[90] Data collated in Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].

Table 2: Results of the WL Archive Copyright Clearance Process

Total Names in Copyright Database	160	Total Identified
Letters requesting permission (sent by WL)	125	
Letters requesting permission (sent by partner archives)	9	
Total Letters Sent	134	84% of those identified
Total Replies	103	77% of those contacted
Permissions received	101	98% of those who replied
Permissions refused	2	2% of those who replied
Did Not Respond	26	19% of those contacted
Low risk – put online after suitable delay	23	
High risk – do not put online	3	
Orphans	22	14% of those identified
Orphans – contact details not found, letters not sent	22	
In Progress	4	4% of those contacted

The success rate for finding contact details for the rightsholders was relatively high: 84%.[91] Staff working on rights clearance for the John Cohen AIDS research collection at the University of Michigan were able to find contact details for 74% of the identified rightsholders (1,023 out of 1,377).[92]However, given the application of the WL's risk criteria, this is not particularly surprising. By focusing on well-known, high-profile and published rightsholders, and controlling their overall numbers through a risk-management strategy, contact details should have been easier to locate.

Of the 134 permission letters sent, 77% (or 101) resulted in permission being granted (98% of those who replied).[93] Digitisation projects that have undertaken an attempt to identify and clear all relevant rightsholders have typically yielded a far lower permission return than this. For example, The National Archive's Moving Here Project sought to obtain permissions to digitise 1,114 wills; they were only able to secure permission for 597 - around half of the selected documents.[94] Another example is provided by a project undertaken by the Koninklijke Bibliotheek of the Netherlands to digitise 1000 Dutch history books. [95] After five months, with a dedicated member of staff for the project, permission for only 50 books had been obtained – 5%

[91] This figure excludes the 14 letters which were not sent to former Wellcome Trust staff – so 22 out of a possible 156 rightsholders details could not be found.

[92] Akmon, D., (2010) "Only with your permission: how rights holders respond (or don't respond) to requests to display archival materials online," *Archival Science,* 45-64, 57.

[93] This percentage is based on the number of permissions granted versus the number of letters sent – 101 permissions from 134 letters.

[94] Vuopala, A., (2010) "Assessment of the Orphan works issue and Costs for Rights Clearance," available at: http://ec.europa.eu/information_society/activities/digital_libraries/doc/reports_orphan/anna_report.pdf [Accessed: 15 August 2013]

[95] *Ibid*

of the overall total.[96] The staff working on the John Cohen AIDS research collection fared marginally better: 68% of the rightsholders contacted for permission to publish responded, (748 out of 1,100 requests), with 94% of those 748 responses granting some form of permission to digitise. However, this still meant that only 64% of the collection (3,490 out of 5,463 items) was made available online, due to the high percentage of non-response to permission requests and orphan works in the collection.[97]

Looking again at the Codebreakers results, in only 2 instances (or 2% of those who replied) was permission to digitise material expressly refused by rightsholders, and in one of these instances, the rightsholder was more concerned with the potential sensitivity of the material, than their intellectual property.

A substantial number of rightsholders (19%) have simply not responded at all. Non-response to requests to digitise material is a real problem with projects of this kind. There are various ways in which a situation of non-response might be explained. The request may have been sent to an incorrect or out-of-date address. Alternatively, the contact details could be correct, but the rightsholder either does not understand the request, or does not care enough about the use of the material in question, to respond. The Wellcome Library legal team expressed a degree of concern about a scenario in which a rightsholder may have received a request for permission to publish, and may care about the use of the material, but does not feel inclined to respond (for whatever reason).[98]

The implication of these scenarios is that archivists should think carefully about how to proceed in situations of non-response. The WL dealt with situations of non-response by reviewing and reassessing the nature of the risk involved in making the relevant material available online without express permission. Where non-respondents were 'known' to the Wellcome (that is: generally, former Trust board members or chairmen, in some cases contacted as a courtesy) then the digitized material was made available. For non-respondents that were not known to the Wellcome, but were nevertheless deemed to be low risk, their work was put online after a further two-month hiatus. This category included scientists and politicians, whose work was regarded as non-commercial, undertaken in the public interest, and therefore the risk of objection was small, although cases had to be considered individually. And finally, for non-respondents that were deemed to be high-risk, the decision was taken not to post anything online without securing express permission first. This category contained the commercial authors and their estates considered the greatest threat of legal action to the project; therefore, permission had to be expressly given before material could be made available online.[99]

It is worth noting that the re-assessed risk criteria the WL used to deal with non-responders could have been used as a guide for how much time to spend searching for addresses, had it been decided upon at the beginning of the project. For example, using the 'low-risk and known to WL' category as a gauge may have resulted in a much shorter search period for those already known to the Wellcome Trust.

The WL's decision to reassess the non-respondents using these risk criteria has allowed them to make an extra tranche of material available online, something which other archival digitisation

---

[96] It is worth noting again that these examples do not include enough detail regarding the number of rightsholder contact details staff were able to locate; nor do they quantify the number of rightsholders who do not respond to permission requests.

[97] Akmon, D. 57

[98] Bird, C. Personal Interview. 18th April 2013.

[99]Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].While three re-assessed risk levels were used during the Codebreakers project, Christy Henshaw has suggested a more flexible approach would be simply to re-assess non-respondents as falling into two categories: "acceptable risk" or "not acceptable risk."

projects have been reluctant to do in similar situations of non-response. Taken together, if we add the lower-risk non-respondents, to those rightsholders who granted permission or could not be located, the total comes to 146 rightsholders. This means that the archive material in which copyright is owned by 91% of the rightsholders identified by the original risk-assessed shortlist will be made available online.

The material related to the 14% of the rightsholders whose contact details could not be found, can be classified as orphan works. Orphan works continue to be problematic for archives; it has been estimated that 21-30% of individual archive collections are made up of orphan works, although in practice it is likely that for individual collections, this percentage will vary greatly, and in most cases be significantly higher.[100] Secondary legislation to address the orphan works problem is expected in the UK during 2014, and, when dealing with archival material, it is likely that some form of diligent search for each individual work selected for publication online will be required.[101] For mass digitization archival projects, and even smaller scale projects, this is clearly not ideal, given the burden comprehensive rights clearance places on archives. Realistically, even when a formal orphan works licensing scheme has been introduced, if the digitization of archival materials is to continue, the risk management strategy adopted by the WL will continue to have relevance across the UK sector.

Finally, it is worth repeating that no fees were paid for the inclusion of any archive material in the project.

4.2.3 Post Project Lessons Learned from Clearing Copyright in Archive Material[102]

Contrasted against the WL's decision not to digitize large amounts of in-copyright books again, the success of the archive clearance effort is unusual for three reasons.

First, it is generally regarded as easier to trace published works than unpublished works, given the commercial interest in published material, and the resources that have been created specifically for tracing rightsholders in published works.

Second, studies have shown that archive collections contain significantly more orphan works than library collections; the process of identifying orphan works in collections is time consuming and requires specific skills training.

Third, as published material contains less sensitive personal data than unpublished material, library collections do not require the high levels sensitivity checking that archive collections are generally subject to, before publication.[103]

The WL's confidence in the archive rights clearance is partly because of its success so far, but is also the result of pragmatism: all of their archive collections are 20th century, so if they didn't attempt some form of clearance, they wouldn't be able to digitise anything at all.[104]

---

[100] Korn, N., In from the Cold: An assessment of the scope of 'Orphan Works' and its impact on the delivery of services to the public (April 2009), available at: http://www.jisc.ac.uk/publications/reports/2009/infromthecold.aspx [Accessed: 25 Aug 2013]

[101] UK Intellectual Property Office, (2013) Orphan Works Licensing Scheme and Extended Collective Licensing FACTSHEET, available at: http://www.ipo.gov.uk/orphanworks-licensing.pdf [Accessed: 30 Sep 2013]

[102] This is a summary of the 'lessons learned' presented in Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].

[103] Deazley, R., and Stobo, V. (2013) Archives and Copyright: Risk and Reform. Working Paper. CREATe / University of Glasgow, Glasgow, UK.

[104] Henshaw, C., Personal Interview. 16th April 2013.

As part of the WL's post-project review of Codebreakers, a number of 'lessons learned' were identified that are of interest. These can be summarised as follows:

(i) When communicating with rightsholders, include as much information as possible about the material held (and preferably include a representative sample), and the part it will play in the project. This is for two reasons. First, the rightsholder is more likely to be enthused by the project and grant permission to digitise. Second, the rightsholder won't have to ask for a clarification of what you hold (which usually relates to concerns over sensitivity issues) which tends to slow down the process of obtaining permission.

(ii) The information provided on the copyright permission and refusal forms should directly reflect that which is available in the relevant policies relating to the project. For example, rightsholders should know that material can only be viewed by registered users, and that they can request to have their material removed from the site at any time. This coherence across the takedown policy, best endeavours search and Access to Archives is essential in clearing articulating what the rightsholder can expect from the project.

(iii) Archive collections will generally include a small proportion of published material such as books, magazines and grey literature. The unpublished materials they contain may have been created in a professional capacity, in addition to the material created personally, and therefore copyright in such material may lie with an employer or other organisation. They also may contain material under Crown Copyright. If you have policies for dealing with these types of material in place at the beginning of the project, digitisers will know whether to digitise the material or not, and cataloguers will be able to record published material in collection descriptions.

(iv) As a process, rights clearance – from identification, tracing, contacting and dealing with enquiries – will take up more time and staff resources than you expect it to. Plan accordingly.

(v) If you rely on the descriptive metadata associated with a collection to gather rightsholder information, double check the original material before sending a request letter: the recipient of a letter can occasionally be mis-identified as the author.

(vi) It is useful to have a single member of staff working on copyright clearance across library and archival material. Occasionally, permission requests overlap where separating the clearance process would have resulted in a rightsholder represented in both library and archive collections, and in collections at the other archives, being contacted two or three times.

(vii) Information about rightsholders and sensitive information in material could be shared between staff working in archives and the content and metadata officers. This raises the possibility that rightsholder information could be captured at the cataloguing stage. For example, if you use personal or organisational names which appear in the collection as a point of access, you are more likely to be capturing information about rightsholders present in the material.

5 Sensitive personal data

The principal focus of the Copyright and Risk project was to evaluate the merits of, and problems encountered by, taking a risk managed approach to copyright clearance. During the course of conducting the interviews which form the basis of this report, it became clear that managing the risks associated with unknowingly publishing sensitive personal data online had been foremost in the minds of the archivists and other staff members involved in the Codebreakers project. The issue of sensitive data was of such concern that it was felt to be worth exploring how the WL and the 3rd party archives managed complying with the Data Protection Act. While the process of sensitivity checking and identifying copyright holders in the archive material was completely separate, the concerns and risks around publication online were judged to be relevant to the project, and worth examining in greater detail.[105]

5.1 Sensitive personal data in context

The Wellcome Library and the partner archives (with the exception of CSHL[106]) comply with both the Data Protection Act (1998) and the Freedom of Information Act (2010),[107] but also observe general ethical principles in the handling of personal information.

Under the terms of the DPA, archives are granted permission to process personal data in their collections under the research exemption (s.33) which provides archives with exemptions from the second and fifth data protection principles,[108] but not from the remaining principles. As the DPA only applies to living individuals, it is standard practice for archivists to apply 100-year closure periods (taking into account the lifespan of the data subject).[109]

It is important to note that the types of sensitive information outlined in the DPA 1998 are not comprehensive; general types of sensitive personal information include:

- Someone's racial or ethnic origin
- His/her political opinions

---

[105] The processes were separate, with different Content and Metadata Officers carrying out different tasks relating to sensitive data and rightsholders. But while some of the policies were separate (Access to Archives does not cover copyright issues) the takedown policy covers both areas: users can object to publication, either as a rightsholder or as a data subject. Rightsholders who were contacted regarding copyright often had queries regarding sensitive data. These areas where copyright and sensitive data appeared to overlap were of interest to the Copyright and Risk project team.

[106] The USA has no equivalent of the DPA; instead they rely upon various laws, regulations and directives which control the use of private data. Two particular acts, The Family Education Rights and Privacy Act (1976) and the Health Insurance Portability and Accountability Act (1996), control the access to medical and student records. The US operates a safe harbour framework which allows US companies to comply with European privacy laws. CSHL comply with New York state law. For Codebreakers, they opted to use the Wellcome's Access to Archives Policy.

[107] The Wellcome Library, as a non-public body, is not subject to FOIA, "but recognises the impact access to the archives of deceased individuals has had both within and outside the public sector." Wellcome Library (Nov. 2010) Access to Archives, available at: http://wellcomelibrary.org/content/documents/access-to-archives.pdf. [Accessed: 12 April 2013].

[108] The second principle states that personal data shall be obtained only for one or more specified and lawful purposes; the fifth principle states that personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes. Data Protection Act, 1998, Sch 1, Part 1.

[109] The National Archives, Society of Archivists, Records Management Society & National Association for Information Management, (2007) Code of Practice for Archivists and Records Managers under Section 51(4) of the Data Protection Act 1998 Available at: http://www.nationalarchives.gov.uk/documents/information-management/dp-code-of-practice.pdf [Accessed: 20 Jul 3013]

- His/her religious beliefs or other beliefs of a similar nature
- His/her trade union membership
- His/her physical or mental health or condition
- His/her sexual life
- The commission or alleged commission of offences by him/her
- Details of any proceedings for any offence committed or alleged to have been committed by him/her, and the outcome of such proceedings including the verdict and, if applicable, the sentence.[110]

But there are other types of information which can be judged to be sensitive,[111] and certain kinds of material, whilst open for physical access in a reading room, can 'acquire' sensitivity if published online.[112] This can include: personal opinions, recommendations or references, both positive and negative, which could have a lasting impact on a data subject's career; expressions of opinion and observations which can be contained in unstructured sources like correspondence or diaries; and administrative sources which contain other personal details (addresses, sponsorships, and so on) which may be deemed to be sensitive. To identify this type of information, the content of the material must be checked, and given the resources that this level of checking would entail, some archives opt to percentage check files.[113]

Given the diverse range of material that was digitised as part of the Codebreakers project, determining whether records contain personal or sensitive data was always going to be a difficult and time-consuming process. Some of the records, such as newspaper clippings and publicity material, could be easily disregarded for the purposes of sensitivity checking as they were already available in the public domain. But with other material, it is much harder to judge: personal correspondence can contain any number of personal details. If a catalogue record provides insufficient detail, the only way to be sure is to check through the original material.

5.2 Dealing with sensitive personal data at the WL

Strict compliance with the legislation would require that each item has to be judged on a case by case basis. However, the Wellcome Library have taken the view that sensitivity checking at item level for all archival material is too resource intensive; instead, they opt to percentage check files.[114] A percentage of a box – say 10% – will be checked for sensitive data, and depending on what is found, the material is either closed, restricted, opened or checked further (perhaps to 50%, or the whole box).

The Wellcome Library has a specific policy, Access to Archives, for dealing with sensitive data held in their collections. This policy predates the Codebreakers project, although it has recently

---

110 TNA Code of Practice available at http://www.nationalarchives.gov.uk/documents/information-management/dp-code-of-practice.pdf [Accessed: 15 August 2013]

111 Browell, G., and Methven, P; Chaplin, S.; Hardy, T.; Richmond, L. Personal Interviews.

112 TNA takedown policy available at http://www.nationalarchives.gov.uk/documents/takedown-and-reclosure-policy.pdf [Accessed: 15 August 2013]

113 Access to Archives available at: http://wellcomelibrary.org/content/documents/access-to-archives.pdf [Accessed: 20 April 2013]

114 *Ibid*

been revised as a result of lessons learned during Codebreakers. It provides guidance that can be followed for all physical, digitised and born-digital material held across the institution, regardless of delivery method or where it is stored. The revisions were necessitated by two particular factors. First, the Wellcome Library uses a remote-ordering system for their on-site reading rooms, which means that access levels for each document must be explicit.[115] And second, as substantial quantities of Wellcome Library material is now being made available online, there has to be consistency between what is available online and what is available in the searchroom, and the various access levels that the Wellcome Library uses.

The sensitivity-checking process for Codebreakers relied entirely on the fact that the collections had already been digitised. A member of the archives team conducted a high-level assessment of the collection using the catalogue entries and the Access to Archives policy, and the result of this assessment was fed to the Content and Metadata Officer with responsibility for Sensitivity Checks. The CMO would then check a percentage of the digital images (original material was not used) and based on the results of this check, edited the metadata to allow the collection to be open, closed or restricted, according to five access categories:

- Open: published works, archives and manuscripts over 100 years old (unless by exception), and non-clinical videos and visual materials such as artworks, photographs, posters, etc.
- Registration Required: un-moderated online registration for archives less than 100 years old with conditions of registration prohibiting users from misusing any private or sensitive data they may find
- Clinical: special permission required by application to access images or video of a clinical nature
- Restricted: no permission to view online but can be consulted in person at the library where archives are considered too sensitive, or where copyright holders have refused permission to publish online (this may be applied to entire items or parts of items)[116]
- Closed: no access allowed to archives at all due to data protection restrictions (but may be digitised for preservation reasons) and born digital content that has not yet been appraised and catalogued or opened to the public[117]

If the original check was inconclusive, the CMO would continue percentage checking until satisfied.

---

[115] For example, at other archives it is a more frequent practice for readers to apply for access to material that is closed or restricted, but this means that an archivist must be available to make a decision on the basis of an application. The Wellcome system is more automated, and therefore needs to be clearly defined.

[116] The Restricted access category is designed to cover scenarios involving both the protection of sensitive personal data and rightsholders wishes regarding online publication. Simon Chaplin has observed that most access considerations are driven by the protection of sensitive data, but that rightsholders may complain about copyright infringement as a proxy for concern about sensitive personal data being made available online. Personal Interview. 18th April 2013.

[117] Henshaw, C., and Kiley, R. The Wellcome Library, Digital Ariadne, Issue 71. Available at: http://www.ariadne.ac.uk/issue71/henshaw-kiley [Accessed: 15 August 2013]

These different access levels presented a number of challenges. The aim of Codebreakers was to provide access to as much material as possible – so parts of some files were closed for sensitivity reasons, but access still had to be available for the rest of the file. The Player uses access information contained in the METS files to prompt for a log-in, which is authenticated using the Wellcome Library's existing Sierra patron database.

5.3. Dealing with sensitive personal data at the 3rd party archives

During the contract negotiations between the WL and the 3rd party archives, it became clear that all of the partners (with the exception of CSHL[118]) wanted some form of reassurance that Wellcome Library would take full responsibility for publishing in-copyright and potentially sensitive material online.[119] None of the archivists involved in the project were particularly worried that they would be sued for breach of copyright, however the potential publication of sensitive data played a much greater role in influencing their behaviour.[120] For all of the external partners, reputational damage proved to be of primary concern,[121] and this concern was directly linked to the potential of inadvertently publishing sensitive personal data online.

Compromises on sensitivity checking had to be negotiated during the project workflow. Each of the UK institutions follows broadly similar internal policies on managing sensitive data, but key differences had to be discussed and resolved. For example, UCL and the WL digitised everything, regardless of whether the material was sensitive or not, relying upon a post-digitisation sensitivity-checking process to determine access levels. However, KCL and GUAS chose not to to digitise any sensitive material.

Cold Spring Harbour Labs had no formal policy in place regarding sensitivity checking, and opted to follow Access to Archives.[122] Ultimately, they adopted a compromise position in relation to their own collection, choosing not to digitise certain parts of Watson's collection which contained medical records.

For one of the partner archives, Churchill, there was no need to engage in any sensitivity checking: the collection had been open to researchers for years as it was judged to contain no sensitive data. However, staff at the WL did re-assess the collection for sensitive data, as the Access to Archives policy explicitly recognizes the difference between making a collection available in a searchroom to individual researchers, and making a collection available online.[123] For the remaining three UK-based partner archives, two divergent approaches were adopted to sensitivity checking.

UCL had a significant amount of cataloguing to complete before digitisation could take place, so a sensitivity check could be carried out simultaneously at item-level. UCL have a similar policy to the Wellcome Library, in that they also have several levels of access, and would normally use percentage checking for collections that have already been catalogued.[124]

---

[118] CSHL demonstrated a more relaxed attitude to copyright infringement than their UK counterparts. Pollock, M. & Zarillo, J., Personal Interview. 30th April 2013.

[119] Chaplin, S., Personal Interview, 18th April 2013.

[120] *Ibid.*

[121] 12 personal interviews with Wellcome and 3rd party archives staff, conducted between 16th April – 7th June 2013.

[122] Pollock, M. & Zarillo, J., Personal Interview. 30th April 2013.

[123] For a detailed description of the risk analysis required see Section 3.3 "Currency of Records" in Access to Archives, page 11. Available at http://wellcomelibrary.org/content/documents/access-to-archives.pdf.

[124] UCL have also pointed out that while the access levels used in Special Collections are bolstered by a robust access policy, percentage checking for sensitive material is a combination of risk assessment and pragmatism, due to the severe constraints of limited staff resources. Gill Furlong of UCL, correspondence with the author, December 2013.

Both GUAS and KCL opted not to digitise material which they found to be sensitive. The collections held by GUAS had been catalogued at the University of Bath,[125] and while KCL already had a comprehensive catalogue available for their collection, the requirements of digitisation meant that both institutions had to create image or item-level metadata for their collections. The process of creating this metadata meant they could more readily assess the collections for sensitive data. Both GUAS and KCL stated that they would not use a percentage checking policy, and both felt it was important that responsibility for sensitive data remained within their respective institutions.[126]

5.4 Conclusion on sensitivity checking

It is worth highlighting that the WL policy for both sensitive data and rights clearance involves a form of risk management – either through controlling the amount of time spent checking material for sensitive data, or controlling the number of rightsholders who must be contacted for permission to publish. The WL were successful in convincing all of the 3rd parties to accept a risk-managed approach to rights clearance, but out of the four institutions which held sensitive personal data in their collections, three out of four did not digitize that material, and in two of those three cases, the 3rd parties rejected the process of percentage-checking files for sensitive data as inappropriate.

This demonstrates that managing sensitive personal data contained within the collections was of greater concern to the 3rd parties and the WL, than infringing copyright law.

[125] The National Cataloguing Unit for the Archives of Contemporary Scientists was established at the University of Bath in 1987, and closed in 2009. The Centre for Scientific Archives at the Science Museum continues to sort, list and index scientific collections of national significance, before finding an appropriate repository for each collection. More details are available at https://sites.google.com/site/centreforscientificarchives/

[126] Richmond, L., Maddra, S., Browell, G., & Methven, P. Personal Interviews.

6 Conclusions

6.1 The success of the WL approach to clearing rights in archive material

Through the narrative of the Codebreakers project related so far, there has been an attempt to systematise the evidence relating to the WL's approach to rights clearance. In particular, the policies used by the WL (that is to say: risk managed rights clearance, a takedown policy, and Access to Archives) have been analysed, alongside the results, whilst the focus on institutional reputations and the way in which the WL have communicated the aims of the Codebreakers project have been examined.

Based on the results of the rights clearance exercise, and the material now available on the WL website, Codebreakers has been a significant success for both the WL and the 3rd party archives involved. In addition to this success, elements of the WL approach to rights clearance could have wider applicability across the UK archive sector.

The WL's risk management strategy had two broad aims: to manage the risks associated with publishing in-copyright material online and publishing potentially sensitive material online. They managed these risks in three ways.

First: the publication of in-copyright material online was made possible using a risk management strategy which involves a set of risk criteria for identifying medium and high-risk rightsholders in the collections, and a diligent search standard for locating and contacting rightsholders.

Second: they managed the publication of potentially sensitive material online using their Access to Archives policy, which provides guidance that can be followed for all physical, digitised and born-digital material held across the Wellcome Library, to take into account both searchroom and online access to sensitive, personal data.

Third: in addition to this, they have an established takedown policy which predates Codebreakers, and applies to all material available on their website.

The most significant indicators of success for the archives clearance process include the number of permissions to publish granted: of the 77% of rightsholders contacted who responded, 98% said yes. The rate of response to permission requests itself was also relatively healthy compared to other projects: the Jon Cohen AIDS collection rights clearance project found that 68% of rightsholders contacted for permission to publish responded,[127] compared to 77% in Codebreakers.

A substantial number of rightsholders (22%) did not respond to permission requests, and although this proportion is lower than in comparable projects, the WL's response is encouraging. By reviewing and reassessing their risk criteria, the WL were able to make much of this material available online, challenging the established practice demonstrated in other archival rights clearance exercises, where non-response forces archives to withhold material.

The material for which 16% of rightsholders contact details could not be found can be classified as orphan works. The Wellcome have made these orphans works available online in batches, but given that the aim of proposed UK orphan works legislation is to make legal use of orphan works possible through diligent search and the payment of up-front licensing fees, and that a diligent search will be required for each individual work, it is inevitable that in order for digitisation of archive material to continue at even a modest scale, risk management of rights clearance is going to have a greater role to play in future. It is hoped that the process tracing outlined in this report, and the example that the WL and the 3rd party archives have provided, will be useful to the sector in this regard.

---

[127] Akmon, D., (2010) "Only with your permission: how rights holders respond (or don't respond) to requests to display archival materials online," *Archival Science,* 45-64, 57.

Money was not an issue when negotiating rights clearance for the archive material; the WL policy was not to offer any kind of reimbursement for the inclusion of material in the project, but in practice, no rightsholders asked for financial payment. One rightsholder requested a donation to a charitable organisation, and the Wellcome decided to donate.

This point is especially pertinent if we consider again that up-front licensing fees are a feature of the proposed UK orphan works legislation. In their Orphan Works Impact Assessment, the IPO estimate that the 'benefits' available to the archives sector for the licensed use of orphan works would lie between £2-76 million p.a., but they state that these benefits are speculative, and that the figures are based on the cost of rights clearance versus the value archives would be expected to derive from the legitimate use of orphan works.[128] However, if rightsholders typically do not request licence fees in relation to the digitisation and publication of their works online, the requirement for archives to pay an up-front licence fee for the use of orphan works is questionable.

The WL were also successful in convincing five 3rd party archives to join the project and accept a risk-managed approach to rights clearance, although this is tempered by the fact that all of the 3rd parties required that some form of indemnity[129] be provided by the WL in the contract negotiations.[130] The greatest area of concern for the 3rd party archives, and the WL itself, was the possibility of sensitive personal data being inadvertently published online.

For a project which set out to examine the rights clearance approach used by the WL, it is clear that managing copyright has been a key factor in the success of Codebreakers, but that managing sensitive data has been of greater importance. This is reflected in the concerns of the staff involved in the project, but also in the rightsholders who contacted the WL in response to permission requests – most of the queries received regarded the sensitivity of the content of the material in question.

Despite the fact that most of the sources consulted during the archive rights clearance exercise were designed with published material in mind, rights clearance for archives has been more successful than the same process for the library material held by the Wellcome Library. For example, for those rightsholders whom the Wellcome were able to find contact details for, only 36% granted permission for the digitisation of library material[131], in contrast to the 75% who granted permission for the archive material. The number of rightsholders who did not respond to permission requests was also lower in the archive material, at 19% in contrast to 28% of rightsholders in library material. The results of the library clearance process show that the digitisation of in-print, in-copyright and published material continues to be problematic for cultural heritage institutions.

Whilst the approach taken in this report has been of a qualitative nature, and given that data is not routinely collected in these kinds of projects, an attempt has been made to systematise the anecdotal nature of much of the evidence gathered. Thanks are due to the Wellcome for the data they collected and shared with CREATe, and it is positive to note that, even when internal processes are not designed for this kind of evidence gathering, it is still possible to gain insights,

---

128 UK Intellectual Property Office, (2012) *Orphan Works Impact Assessment*, p.11. The Impact Assessment also values the indirect commercial value of orphan works at £9-91 million p.a. Available at: http://www.ipo.gov.uk/consult-ia-bis1063-20120702.pdf [Accessed: 30 Sep 2013].

129 Chaplin, S. Personal Interview, 18th April 2013.

130 "The Wellcome Library agreed to be liable for copyright infringement for works published on the Wellcome Library site, while the 3rd party archives took responsibility for assessing sensitivity and telling the Library what to suppress. The Wellcome Library do not indemnify the 3rd party archives for any content they display on their own websites, or share with others. They do this at their own risk." Christy Henshaw, correspondence with the authors, December 2013.

131 Rightsholders were contacted in relation to 1365 titles. This is made up of permissions received, permissions denied, outstanding queries and those that did not respond.

data and evidence to help inform the debate. At a *Copyright and Risk* dissemination event organised on 27th September 2013 and held at the Wellcome Trust[132], the IPO made very clear that they consider this kind of evidence to be immensely valuable, and if we can collect and present it consistently, they are extremely willing to engage with the sector on issues that concern archivists. Section 6.5 of this chapter lists some of the evidence which would be particularly useful for these purposes.

In the next sections, we present some of the elements of the WL approach which we think are of greatest relevance and use to the UK archive sector.

6.2 The Policies

6.2.1 The Takedown Policy

The Wellcome Library takedown policy is an exercise in risk mitigation: it allows the Wellcome a measure of control should a rightsholder or data subject object to the publication of material on the WL site, and also allows them to demonstrate that they acted in good faith on receipt of such a complaint. The policy is designed to be quick response: the material is removed immediately on receipt of a takedown request, and a review panel assesses the requests on a case by case basis, returning a decision within four weeks.

The takedown policy has also been useful when contacting rightsholders for permission to publish their works online: because the WL have tried to avoid short-term licensing agreements, they can use the takedown policy to reassure the rightsholders that they can request that their material is removed from the WL website at any time, regardless of whether permission was given for publication in the past.

It should be standard practice for any type of online, publicly accessible repository to publish a takedown policy on their website: it's the simplest step in the risk mitigation process, applies to all material published on the site, and can provide some protection from complaints resulting from the publication of copyright works, sensitive personal data and obscene or defamatory material.[133] Several of the partner archives also had takedown policies available on their websites before they started working on the Codebreakers project.[134]

Since the launch of Codebreakers in March 2013, and in relation to 1.6 million digital images, the WL have received only one takedown request for the digitised archive material that has been published on the website, and this request appears to have been prompted by the content of the material, not its copyright status.

6.2.2 Risk Management

The risk management strategy used by the WL covers several points which are essential for other archives to consider when thinking of undertaking a digitisation project which may include both copyright-protected material and sensitive data.

---

[132] *Archives and Copyright: Developing an Agenda for Reform,* Wellcome Trust, London, 27th September 2013. Joint conference organised by CREATe and supported by the Wellcome Trust and the Archives and Records Association of the UK and Ireland.

[133] For example, the JORUM takedown policy states that material "which is unlawful e.g. breaches copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, or which otherwise constitutes a breach of the Jorum Terms of Service," can be removed. JORUM Notice & Take Down Policy, available at: http://www.jorum.ac.uk/policies/jorum-notice-and-takedown-policy [Accessed: 1 Sep 2013]

[134] Browell, G., and Methven, P.; Pollock, L., and Zarillo, J.; Furlong, G., Makin, K., and Wright, S. Personal Interviews.

Firstly, the WL state clearly on their website that they undertake risk assessments for both copyright material and sensitive data – where possible – and in addition have taken the further steps to reduce risk by:

- Requiring users of the Wellcome website to either agree to terms and condition of use, or to register with the library, to view some types of material
- Only making material available under non-commercial licences (unless other licence agreed by rightsholder) and where required, blocking download
- Restricted online access to material under 10 years old.[135]

Requiring that users register to view most of the material available on the website obscures some of the risk that the WL have taken on – it's not difficult to believe that a casual user of the site will navigate away from the Player when asked to register. By providing different sign-in options, users are offered more convenience, but the registration requirement runs against the otherwise open-access aims of the project. This is further reinforced by the decision to restrict online access to material less than 10 years old, regardless of whether the material contains sensitive information or not.

In addition to the takedown policy discussed previously, the WL also provides a statement which outlines how copyright was cleared in the digitised collections. They explain that they have worked with ACLS and PLS to clear copyright in print material where possible, and for archive material, "we have made all reasonable efforts to ensure copyright holders' interests are respected and permission sought where it is feasible to do so. However, we cannot guarantee to have traced or contacted every potential rights-holder."[136] Statements like this, which explain the difficulty of the rights clearance process, are useful not only in terms of risk mitigation, but also for users who may be considering re-using or re-publishing a work in some form. This type of statement can also be extended to solicit information about collections which have been made available online, and this is particularly useful for collections where rightsholder information is scarce or unavailable.[137]

Given their previous experience, and the scale of the project, there was never an intention of comprehensive rights clearance on the WL's part. They were open to collaboration, and their risk criteria were further augmented by the concerns of the 3rd party partner archives involved in the project. Definitions of risk are fluid, and should be adapted to the organisations and projects at hand. An archive choosing to digitise and make copyrighted works freely available on the internet must gauge their own risk appetite and establish their own risk criteria prior to embarking on the project.

With risk defined, the WL was able to identify rightsholders through indexes, catalogues and administrative records. Record-keeping and cataloguing practices within archives could be used to make identifying rightsholders a simpler process, although in practice this is very time-consuming as each copyright holder would need to be listed in the catalogue, metadata or

---

[135] The three steps are taken from the 'Guiding Principles' section of the Wellcome Library's Copyright notice and takedown statement available at: http://wellcomelibrary.org/about-this-site/copyright-clearance-and-takedown [Accessed: 1 August 2013] It is worth noting that the first and last of these steps apply only in consideration of sensitive data, not copyright.

[136] Ibid

[137] Hirtle, Hudson and Kenyon also encourage solicitation about collections in order to placate unidentified copyright owners – "The dialogue may also begin a mutually beneficial discussion of permission, licensing, or takedown." Pg. 208, Hirtle, P.B., Hudson, E., and Kenyon, A.T., (2009) *Copyright & Cultural Institutions: Guidelines for Digitisation for US Libraries, Archives and Museums*, Cornell University Library, Ithaca, New York.

records. Although identifying the names would be easier, there would also be significantly more names to assess, whereas (in the Wellcome's case) names "hidden" in the content, and not available in the catalogues or metadata were essentially invisible and therefore were not considered for copyright clearance.

When negotiating deposit agreements (where possible) archivists should argue to have creator copyright in the collection transferred to the archive – especially as more archives are accessioning born digital material. Obviously this is not without complications; depositors may not own the copyright in all or any of the material they are depositing, but having robust policies in place can help you to persuade rightsholders to sign such permissions over to the archive.

The WL used a variety of sources to trace and contact rightsholders, and through keeping records of this process, have demonstrated a practical and auditable[138] approach to due diligence.

6.2.3 Access to Archives[139]

Given that protecting sensitive information within archives is so important to both archivists and rightsholders,[140] having a robust and defendable policy in place for handling sensitive data is advisable. The policy should be applicable to the actions of the institution regardless of whether viewing and copying material occurs online or in the searchroom.

Issues over sensitivity were also where the 3rd party partners have shown the most divergence, and the policies and attitudes at Wellcome and UCL differ to those at KCL and GUAS.

Much emphasis in the literature has been placed upon the cost, time and effort involved in rights clearance, which does not include costings for, or mention of, the equally arduous task of checking material for sensitive information, and the judgement required to perform such a task. A lack of quality and qualitative literature on the types of sensitive data held in UK archives, and of examples of archivists dealing with such data, was noted while preparing the Working Paper, *Archives & Copyright: Risk and Reform*.

6.3 Reputation

The WL, and all of the 3rd party archives involved, cited reputational damage as the most risky element in running and taking part in Codebreakers.

Reputational damage is linked to specific concerns:

- That sensitive information about a data subject might be published online
- That a rightsholder may object to the use of their work
- Those institutional policies may be exposed as insufficiently robust, or that taking part in the project may be interpreted as inconsistent with the previous behaviour of the archive.[141]

The 3rd party partner archives expressed existing relationships with depositors, and attracting future depositors, as being of primary concern.[142] Archivists at the Wellcome Library also

---

[138] Henshaw, C., Personal Interview. 18th April 2013. Christy believes there may be potential for auditing the diligent search practices of cultural heritage institutions.

[139] The Wellcome Library Access to Archives policy is available at: http://wellcomelibrary.org/content/documents/access-to-archives.pdf

[140] Both Christy Henshaw and Caroline Herbert mentioned that rightsholders often wanted to see copies of the material they were being asked to give permission to publish - this was usually casued by concern at the contents, not their copyright status. Henshaw, C.; Herbert, C. Personal Interviews.

[141] Packwood, A., Adams, N., and Riley, A. Personal Interview. 9th May 2013.

differentiated between complying with data protection and copyright legislation in that data protection legislation and the associated ethical considerations of personal data are taken more seriously than copyright legislation.

The reputation of the Wellcome Trust, and by extension, the Library,[143] also had a role to play in the success of the project, and specific rightsholders were described as 'delighted' to be included in Codebreakers.[144] Communicating, articulating and maintaining the reputation of the Wellcome Trust, and by extension the Wellcome Library as a trusted, reliable repository, is another successful facet of the *Codebreakers* project and the WL's approach. It is present in the projects they have engaged in, the formulation of their policies, and how they communicate with stakeholders, rightsholders and users. Making material available in the way the WL have could be seen as gambling with that reputation in the long term, but they also have a commitment to open access to uphold.

6.4 Communication

Articulating and communicating the aims and objectives of the digitisation project engaged in is of crucial importance in winning over rightsholders. The WL found that the more information they included in the initial letter – the more likely they were to get permission granted. This allayed rightsholders primary concerns – which were generally around the content of the material, rather than their ownership. Conversely, the WL found that making the permission and refusal forms as simple and as clear as possible – and having robust policies to link them to – was also crucial in making it as easy as possible for rightsholders to respond.[145] For similar reasons, email (when discoverable) was favoured over posting letters to secure permission.

A member of staff at one of the 3rd party archives gave an example of tracing and contacting the heirs of a primary rightsholder in one of the collections selected for digitisation.[146] Making the initial approach to a primary rightsholder, or their relatives, can be a fraught process. Will the rightsholder expect to be paid for material that was (in the vast majority of cases) created with no expectation of financial gain? Will they only grant permission for a temporary license, meaning the whole process will start again in five or ten years' time? What is their relationship to the other members of their family, what relationship did they have with the primary rightsholder, and how do they feel about the material in question?

Again, answering these types of questions, and negotiating the terms of permission, is easier when you have a coherent story about the project, and your archive, to tell the rightsholder.

Various access levels, download restrictions, licences, and the awkward graft of the archive hierarchies into the Wellcome Library's main Encore catalogue (designed to display library catalogues, not archival catalogues) are communication and user issues the WL will have to address in the long-term. While a dedicated researcher will persevere regardless, the 'curious public'[147] that the Wellcome Library are keen to target may find the sign-in requirement off-putting, and the navigation within the collections and identification of digitised content confusing.

---

142 Browell, G., and Methven, P. Personal Interview. 28th May 2013.

143 A global charitable foundation, dedicated to achieving extraordinary improvements in human and animal health, with one of the biggest resources for medical history in the world, founded by Henry Wellcome and opened in 1945. http://www.wellcome.ac.uk/

144 Maddra, S. Personal Interview. 2nd May 2013.

145 Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].

146 Maddra, S. Personal Interview. 2nd May 2013.

147 The tagline for the Wellcome Collection is "A free destination for the incurably curious." See http://www.wellcomecollection.org for more information.

6.5 Evidence to collect about archival rights clearance exercises

If an archive service intends to engage in rights clearance for a digitisation project, there are certain types of information it would be useful to gather throughout the project. This information is useful at an institutional level: it can be used to determine the success of the project, and to highlight considerations for future projects. Such evidence will also be useful to the sector as a whole, if it can be collected and presented systematically to policy makers and legislators, in order to help them better understand how the copyright regime impacts upon the work of archivists and the institutions they work for.

Information can include:

- The risk criteria used (if any) by the institution, when deciding which rightsholders to contact
- The agreed limits of diligent search at a particular institution
- The length of time taken to conduct a diligent search
- The financial cost of diligent search in particular, and rights clearance processes in general
- The number of permissions granted by rightsholders
- The number of refusals by rightsholders, with reasons for the refusal recorded where possible
- What level of reimbursement (if any) was expected in return for permission to publish, and whether that reimbursement was given
- The number of rightsholders who failed to respond to permission requests
- The number of rightsholders for whom contact details could not be found
- The number of orphan works in individual collections

7 Appendixes
7.1 Example of an Archives Permission Letter

Copyright Owner's Name
Copyright Owner's Address
Date:

Dear Name

FOUNDATIONS OF MODERN GENETICS – Request for Copyright Clearance

The Wellcome Library specialises in the history of medicine and holds books, manuscripts, archives, films and pictures from the earliest times to the present day. The library is free and open to the public. The Wellcome Library is part of the Wellcome Trust, a global charitable foundation dedicated to improving human and animal health. The Wellcome Trust is independent of both political and commercial interests.

The Wellcome Library is working on a large-scale digitisation programme to improve access to its collections. The 'Foundations of Modern Genetics' is a central theme of this digitisation programme and we are aiming to make material on this history of genetics in the 20th century available on our website. To realise this aim, we will be digitising content from the archival and printed collections of the Wellcome Library and its five partner institutions (Cold Spring Harbor, New York; Churchill Archives Centre, Cambridge; Archive Services, University of Glasgow; King's College London and University College London). We plan to launch in late 2012.

Among the documents that have been identified for inclusion in this project are [describe the material, e.g. the letters written by Dr. Smith while he was working at XXXX in 1940s-1970s which are held by Archive Services, University of Glasgow]. We believe you may be the current copyright holder, and we would like your permission to digitise this material and make it accessible online, as it would be a vital contribution to the project.

These important collections will be freely available to the worldwide community of researchers and the general public, and will document the astonishing development of genetic science over the last century. We are aware that some of this material may contain sensitive personal information and therefore we have developed a robust policy to ensure that it will be treated in a responsible and lawful manner. You can view our Access to Archives policy on http://library.wellcome.ac.uk/doc_wtx058082.html. We will not be charging for access to this material, and it is not the intention of the Wellcome Library, nor will it ever be, to exploit this material for profit. Any use of the material will be restricted to non-commercial purposes.

We would be delighted if you could complete and sign the Copyright Clearance Form (in Annex A). If you do not wish the Wellcome Library to use the documents, or you are not the copyright holder, then please return the Copyright Refusal Form (Annex B) to us. We have enclosed a stamped, addressed envelope to return the forms to the Wellcome Library.

We have also included a list of "Frequently Asked Questions" (in Annex C) to provide you with further information relating to the WL. If you have any queries, please contact Caroline Herbert at c.herbert@wellcome.ac.uk.

Yours Sincerely,

Annex A
Copyright Clearance Form

The Wellcome Library is delighted that you will support our commitment to providing online public access to historic content. Please complete the following:

[Insert name of individual or organisation] hereby grants permission to the Wellcome Trust to:

·Digitise the following works: Describe works here

·Make the digitised works listed above available online, provided that for any reuse, adaptation and distribution, users must abide by the Wellcome Library's Conditions of Use, which will provide a licence to use the works for non-commercial purposes.

Signature: _____

Print Name: _____
On behalf of [insert name of person or organisation]

Title/Position: _____

Date: _____

Annex B
Copyright Refusal Form

If you do not wish to allow the works to be digitised by the Wellcome Trust, please complete the following:

I am the copyright holder of [describe the material] and I hereby decline permission for the Wellcome Trust to digitise and make these materials accessible online.

It was be useful for us to know why permission has been refused. Please use the space below to explain.

Signature:_____

Print Name:_____

Title/Position: _____

Date: _____

Or, if you are not the copyright holder, please check this box:

If you know who the copyright holder is, please provide details below:

Name:	
Address:	

Reference:

Annex C
Frequently Asked Questions

What is the Wellcome Library?
The Wellcome Library, part of the Wellcome Trust, specialises in the history of medicine and holds books, manuscripts, archives, films and pictures from the earliest times to the present day. The library is free and open to the public. The Wellcome Library is part of the Wellcome Trust, a global charitable foundation dedicated to improving human and animal health. The Wellcome Trust is independent of both political and commercial interests.

What is the Wellcome Digital Library?
This is a new online resource that will be available to everyone free of charge from Spring 2013. The initial launch will provide access to a pilot digitisation project based on the theme of the 'Foundations of Modern Genetics'.

Who are the five archives and libraries who are contributing material?
Cold Spring Harbor, New York; Churchill Archives Centre, Cambridge; Archive Services, University of Glasgow; King's College London and University College London.

What material is being digitised for the Wellcome Digital Library?
For the pilot, a mixture of books, archives and images around the theme of 'Modern Genetics' will be digitised. A list of material being digitised from the Wellcome Library's collections can be found at: http://library.wellcome.ac.uk/doc_wtx058366.html.

How will the public access the Wellcome Digital Library?
The public will be able to view digital images of the archives on their computer screens, download images and PDFs, and share images according to Conditions of Use.

Will users need to register to use the Wellcome Digital Library?
Many types of material will be available without registering, but to view archival material that is less than 100 years old we will require users to register online.

What will happen to the original archives?
All the original papers, photographs etc. will be preserved by the holding institutions and will continue to be available for researchers to use. Digitisation will allow more people to use the material remotely and avoid damaging the originals.

Under what terms will people be allowed to use the material on the Wellcome Digital Library?
Users will be able to use material under the terms of a non-commercial licence and registered users will have to abide by data protection and copyright law. For example, a teacher could use copies of a letter printed out from the Wellcome Digital Library in a presentation to students, as long as they had undertaken to abide by the WL's published Conditions of Use.

What about commercial use?
Any requests to use the material for commercial purposes will be directed to the institution holding the original documents, who can allow commercial use under the usual terms for that institution's reprographics services.

Will the Wellcome Library make money from this project?
No. The Wellcome Library will not sell images of these works or require subscription fees or any other type of payment from the public with the exception of providing high quality images for publications or other professional use. In these exceptional cases an administration fee will be charged.

Are creators/authors/publishers paid for granting permission to use their material on the WL?
No. The aim of the Wellcome Digital Library is to make available copies of important archive collections which are already preserved by, and freely available to see at, the partner organisations. This project is being undertaken to promote research and interest in the history of genetics for the benefit of all. We hope that copyright-holders will support this goal by allowing material to be digitised without seeking payment.

How will the Wellcome Digital Library deal with sensitive personal data?
Some of the collections being digitised contain sensitive information, e.g. medical records and other personal information. The Wellcome Library's Access to Archives policy outlines the steps we take to balance the Library's duty of care to individuals mentioned in collections with the legitimate requirements of researchers, and can be found at: http://library.wellcome.ac.uk/about-this-site/copyright-clearance-and-takedown/. All archive collections are assessed for sensitive data before they are made available online.

What security measures are in place to protect the digital copies of my works held by the Wellcome Library?
The Wellcome Library has a secure digital library system that can only be accessed by authorised users. Images are permanently backed up to an off-site storage area for long-term preservation.

What will happen to the Wellcome Digital Library in the future?
This new and exciting initiative will be the foundation of a global resource for the history of medicine and health. The current project on genetics is a pilot but we expect to build on our digitisation work in the future, expanding into areas such as public health, mental health, environment, and infectious disease.

7.2 Example of a Library Permission Letter

Copyright Owner's Name:
Copyright Owner's Address
Date:

Dear Sir/Madam,

Re: Seeking permission to digitise your in-copyright book(s)

The Wellcome Library is committed to the long-term preservation of the books in its collections, to ensure they can be accessed by researchers for generations to come. To further this aim the Wellcome Digital Library will be launched in 2012 with a range of archives and books related to the history of genetics, genomics and inheritance. This new and exciting initiative will be the foundation of a global resource for the history of medicine and health.

This first collection comprises approximately 2000 books published between 1850 and 1990 from 53 different countries. As the majority of the titles in this collection are in copyright (but out of print), identifying and tracing the authors and publishers to seek permission to use this content presents a major challenge.

The Authors Licensing and Collecting Society (ALCS) and the Publishers Licensing Society (PLS) have contributed to the development of ARROW, a system that will deploy their respective works databases and networks of international rights-holders to provide a rights-identification search and contact service for the authors and publishers of the in-copyright books. The Wellcome Library is using ARROW to seek out the relevant copyright owners.

According to research, carried out through ARROW, we have determined that you may own the copyright to some of the title the Library would like to include in the project, as listed in Annex A. For each work, please confirm whether or not you own the rights. In those cases where you do hold the rights, please indicate whether you are willing to give permission to the Wellcome Library to digitise these works, and if so, under what licence terms.

The Wellcome Library requests permission to digitise these works and make them available online in the Wellcome Digital Library on terms that will allow the public to freely access this content for activities such as private study, academic research and teaching. Commercial use of this content will NOT be permitted. Further information about possible uses (and what uses would not be permitted) are outlined in the Frequently Asked Questions leaflet, enclosed as Annex C.

If you are happy for your works to be included in this innovative digitisation project please complete the attached spreadsheet, sign the enclosed Copyright Clearance Form in Annex B, and return to the Wellcome Library using the enclosed SAE. Please note that copyright owners granting the necessary licence to Wellcome will be able to request that their material be removed from the Wellcome Library website at any time and for any reason.

If you do not wish to include your works in this innovative digitisation project or are not the copyright holder, but you know who the copyright holder is, please complete and sign the

enclosed Copyright Refusal Form in Annex B and return to the Wellcome Library using the enclosed SAE.

Any queries can be directed to the Wellcome Library at: [insert email address]

Yours Faithfully

Simon Chaplin
Head of the Wellcome Library

Annex A
Your titles within the collection

Title	Author	Do you own the rights to this work? [Yes/No]	In cases where you own the rights, are you willing to give permission to the Wellcome Library to digitise this work and make it available over the internet? See Annex B for details. [Yes/No/Not applicable]	Licence Restrictions: [None; Restriction A; Restriction B; Restriction C]

Annex B: Copyright Clearance Form

The Wellcome Library is delighted that you would like to participate in the Wellcome Digital Library (WDL) to further its commitment to the long-term preservation of the books in its collection to ensure that they can be accessed by researchers for generations to come. Please complete the following:

I, (name) of (address)

Hereby grant to the Wellcome Trust to the extent of my ownership or control of the rights in works listed in Annex A ("works") a licence to enable the Wellcome Trust to:

- Digitise the works; and
- Make the works available to users of the WDL provided that:
  - -The user fully attributes the work in any manner specified by below;
  - -The work may not be used or distributed for commercial purposes whether by means of the sale, resale, loan, transfer, hire or any other form of exploitation of the works;
  - -For any reuse or distribution of the work the user must make clear to others the terms of this licence; and
  - -This licence is subject to any restriction specified below.

For the avoidance of doubt the following rights are not affected by the licence:

- The user's fair dealing or fair use rights, or other applicable copyright exceptions and limitations;
- The author's moral rights; and
- Any rights other persons may have either in the work itself or in how the work is used, such as publicity or privacy rights.

I may notify the Wellcome Trust at any time and for any reason to remove any of the works from the WDL and the Wellcome Trust will block access to the notified works and will not display them without my further permission.

Restrictions

Please indicate below, by marking the relevant boxes, if you wish to prevent the user from:

- Copying by means of downloading the whole work [Restriction A]
- Sharing (issuing, performing or communicating) copies of the work in whole or in part (in print or electronic form) with others [Restriction B]
- Altering, adapting, modifying or translating the work [Restriction C]

All terms used are as defined in the Copyright, Designs and Patents Act 1988.

Attribution

In all cases the user must fully attribute the copyright owner of the work as follows "(c) [copyright owner name] [date of publication]" or if you wish to specify a different attribution, in the manner specified below:

Please provide attribution notice:

Signature Date

Annex B: Copyright Refusal Form

If you do not wish to participate in the Wellcome Digital Library, we should be grateful if you could complete the following:

I, (name) of (address),

Do not give my permission to the Wellcome Digital Library to digitise the works listed in Annex A ("works").

Please provide your reasons for refusing permission in the box below:

Signature Date

If you are not the copyright holder, but know who the copyright holder is, please provide details below:

Name	
Address	

Text of Letter sent to Rightsholders (3rd Letter)

Dear Sir/Madam,

Re: Seeking permission to digitise your in-copyright book(s)

The Wellcome Library has written to you about a book or books for which we believe you are the copyright-owner. A copy of our most recent letter is attached.
Although we have had a positive response from many authors and publishers, there are a large number of books – including yours in the attached list – for which we have had no reply. We would like to make these books available online from 1st February 2013, as part of our on-going project to improve understanding of the history of genetics for researchers and the public. The books will be available for non-commercial use only.

The Wellcome Library is part of the Wellcome Trust, a global charitable foundation which supports research in the biomedical sciences and the medical humanities. The Wellcome Library will not receive any financial benefit from making the works available: our aim is simply to support those interested in the history of biomedical science to access published research that is no longer available in print.

If you do not want the books listed overleaf to be made available online, or would like us to apply additional restrictions on their use, please let us know using the form provided. If you are not the copyright owner, please tell us as we will remove your details from our database.

Any queries can be directed to the Wellcome Library at: [insert email address]

Yours Sincerely,

Simon Chaplin

Head of the Wellcome Library

7.3 WL Archive Risk Criteria

Guidance on Assessing Copyright Risk for Archives

For the archives destined to be part of the WL we are taking a risk managed approach to copyright. This means that we will not be contacting all the copyright holders. We will only contact those who appear to present a risk. Permission from the named creator / donor of a collection, e.g. Watson or Pontecorvo, should already be in place. This process concerns material by other copyright holders held within the named collections. The first step in this process is to identify the higher risk copyright holders.

Creating a shared list of potentially risky copyright holders

Every institution contributing material to the WL will use the metadata to identify the copyright holders who present a risk. If the name does not appear in the catalogue record, then they are not considered a risk. These names should be added to the shared list on www.dropbox.com . The list is only for material that is due to be included on the WL. If a decision has already been taken to exclude material in a collection there is no need to put details on the list. Assessing risk is not an exact science. We need to make a judgement based on a number of factors and how these factors combine. The bullet points below are designed to help you to decide whether or not to include a copyright holder on the long list. At this stage include any names you are unsure of. We can always remove them at a later stage.
Qualifying criterion - For a name to be included on the long list it must appear in the metadata / catalogue records.

Medium Risk – When all of these three factors apply the material is classed as medium risk and should be added to the long list,

- The author / creator has (or had) a high public profile.
- The author / creator is alive or is known to have a literary estate (as recorded in http://tyler.hrc.utexas.edu//)
- The material appears to have been published / broadcast and / or prepared for commercial gain, rather than to advance academic knowledge or in a not-for-profit environment.

High Risk – If any one of the following apply the material is classed as high risk and should be added to the long list,

- The author / creator is a well-known literary figure, broadcaster or artist.
- The author / creator / literary estate / publisher is known to actively defend their copyright
- The relationship between the holding institution and the author / creator / publisher is awkward.
- There is a large proportion of material from an author / creator which appears to be in copyright, i.e. more than 20% in a box.

[Document dated 9th November 2011]

8 Bibliography

Adams, N, Packwood, A, & Riley, A. Personal Interview. 9th May 2013.

Akmon, D., (2010) "Only with your permission: how rights holders respond (or don't respond) to requests to display archival materials online," Archival Science, 45-64

Bird, C. Personal Interview. 18th April 2013.

Browell, G. & Methven, P. Personal Interview. 28th May 2013.

Copyright, Designs and Patents Act 1988 (UK)

Chaplin, S., Personal Interview, 18th April 2013.

Covey, D.T., (2005) 'Response to Notice of Enquiry about Orphan Works,' Carnegie Mellon University Libraries. Available at: http://www.copyright.gov/orphan/comments/OW0537-CarnegieMellon.pdf [Accessed: 15 August 2013]

Covey, D.T., (2005) *Acquiring copyright permission to digitise and provide open access to books*, DLF, Council on Library and Information Resources, Washington DC. Available at http://www.clir.org/pubs/reports/pub134/reports/pub134/pub134col.pdf [Accessed: 20 Sep 2013].

Deazley, R., and Stobo, V. (2013) *Archives and Copyright: Risk and Reform*, CREATe Working Paper No. 3, University of Glasgow, Glasgow, UK. Available at: http://www.create.ac.uk/wp-content/uploads/2013/04/CREATe-Working-Paper-No-3-v1-1.pdf [Accessed: 20 April 2013]

Dryden, J., "Copyright issues in the selection of archival material for internet access" (2008) Archival Science 123-47

Furlong, G, Makin, K, & Wright, S. Personal Interview. 6th June 2013.

Hardy, T., Personal Interview. 17 April 2013.

Henshaw, C and Herbert, C. (2013) *Wellcome Library digitisation pilot project: notes on copyright clearance*, Wellcome Library Internal Document [Accessed: 15 April 2013].

Henshaw, C. & Kiley, R., "The Wellcome Library, Digital". July 2013, Ariadne Issue 71. Available at: http://www.ariadne.ac.uk/issue71/henshaw-kiley [Accessed: 20 July 2013].

Henshaw, C. "Managing Large Scale Digitisation at the Wellcome Library" (presentation, Sync or Sink: Opportunities for Libraries in the Digital Age, Birkbeck College, University of London,

24 Nov 2011) Slides available at: http://www.slideshare.net/Wellcome/cdp25-nov2011 [Accessed: 13 July 2013]

Henshaw, C. "Workflow Thinking for Digitisation," (presentation, Digitisation Doctor, Wellcome Trust, London, 15 April 2013). Slides available at: http://prezi.com/unqxvaz5swvp/workflow-thinking-for-digitisation/ [Accessed: 1 July 2013]

Henshaw, C. in Deazley, R., and Stobo, V., (eds) (2014) *Archives and Copyright: Developing an Agenda for Reform*, Digital proceedings of RCUK symposium [Forthcoming]

Henshaw, C., Personal Interview. 16th April 2013.

Herbert, C. in Deazley, R., and Stobo, V., (eds) (2014) *Archives and Copyright: Developing an Agenda for Reform*, Digital proceedings of RCUK symposium [Forthcoming]

Herbert, C., Personal Interview. 16th April 2013.

Hirtle, P.B., Hudson, E., and Kenyon, A.T., (2009) Copyright & Cultural Institutions: Guidelines for Digitisation for US Libraries, Archives and Museums, Cornell University Library, Ithaca, New York.

Hubbard, B., (2008) "Take Down Policies" JISC Repositories Support Project, available at: http://www.sherpa.ac.uk/documents/take-down.pdf [Accessed: 2 September 2013]

JORUM Notice & Take Down Policy, available at: http://www.jorum.ac.uk/policies/jorum-notice-and-takedown-policy [Accessed: 1 Sep 2013]

Korn, N., (2009) "In from the Cold" JISC & The Collections Trust, available at: http://www.jisc.ac.uk/publications/reports/2009/infromthecold.aspx [Accessed: 1 Sep 2013]

Maddra, S. Personal Interview. 2nd May 2013.

Packwood, A., Adams, N., and Riley, A. Personal Interview. 9th May 2013.

Pollock, M. & Zarillo, J., Personal Interview. 30th April 2013.

Richmond, L., Personal Interview. 26th April 2013.

Stratton, B., (2011) "Seeking New Landscapes: A rights clearance study in the context of mass digitisation of 140 books published between 1870 and 2010," British Library, London, UK. Available at: http://pressandpolicy.bl.uk/imagelibrary/downloadMedia.ashx?MediaDetailsID=1197 [Accessed: 1 June 2013]

The National Archives Code of Practice available at http://www.nationalarchives.gov.uk/documents/information-management/dp-code-of-practice.pdf [Accessed: 15 August 2013]

The National Archives takedown policy available at http://www.nationalarchives.gov.uk/documents/takedown-and-reclosure-policy.pdf [Accessed: 15 August 2013]

The National Archives, Society of Archivists, Records Management Society & National Association for Information Management, (2007) Code of Practice for Archivists and Records Managers under Section 51(4) of the Data Protection Act 1998 Available at: http://www.nationalarchives.gov.uk/documents/information-management/dp-code-of-practice.pdf [Accessed: 20 Jul 3013]

UK Intellectual Property Office, (2012) Orphan Works Impact Assessment, p.11. Available at: http://www.ipo.gov.uk/consult-ia-bis1063-20120702.pdf [Accessed: 30 Sep 2013].

UK Intellectual Property Office, (2013) Orphan Works Licensing Scheme and Extended Collective Licensing FACTSHEET, available at: http://www.ipo.gov.uk/orphanworks-licensing.pdf [Accessed: 30 Sep 2013]

Vuopala, A., (2010) "Assessment of the Orphan works issue and Costs for Rights Clearance," available at: http://ec.europa.eu/information_society/activities/digital_libraries/doc/reports_orphan/anna_report.pdf

Wellcome Library (2010) *Access to Archives*, available at: http://wellcomelibrary.org/content/documents/access-to-archives.pdf. [Accessed: 12 April 2013].

Wellcome Library Terms and Conditions, available at: http://wellcomelibrary.org/about-this-site/terms-and-conditions/ [Accessed: 1 September 2013]

Wellcome Library, (2008) *AIDS Campaign Posters – Copyright management strategy*, Internal Document [Accessed: 23 April 2013]

Wellcome Library, (2011) *Guidance on Assessing Copyright Risk for Archives*, Internal Document [Accessed 20 April 2013] Available in the Appendix.

Wellcome Library, Collection Development Policy is available at http://wellcomelibrary.org/content/documents/policy-documents/collection-development-policy.pdf [Accessed: 5 Dec 2013]

Wellcome Library, Copyright notice and takedown statement available at: http://wellcomelibrary.org/about-this-site/copyright-clearance-and-takedown [Accessed: 1 August 2013]

Wellcome Library, Takedown Policy is available at: http://wellcomelibrary.org/about-this-site/copyright-clearance-and-takedown [Accessed: 01 Aug 2013]

Wellcome Library, *Transforming the Wellcome Library 2009-2014*, available at: http://wellcomelibrary.org/about-us/library-strategy-and-policy/transforming-the-wellcome-library/ [Accessed: 30 Sep 2013]