

DIGITISING THE EDWIN MORGAN SCRAPBOOKS: DIGITISATION AND RISK

Victoria Stobo, Kerry Patterson and Ronan Deazley

1. INTRODUCTION

Decisions about copyright clearance, when to do it, how to do it and how much to do, are always considerations based in the end on a vision of risk, and of risk tolerance in a particular institution ...

Professor Peter Jaszi

Curators, archivists and librarians have to balance a series of demands, expectations and risks when they make digitised collections available online. Before the introduction of the EU Directive and the development of the UK's Orphan Works Licensing Scheme (OWLS), cultural heritage institutions (CHIs) had to determine their own approach to using orphan works. Sectoral guidance included a recommendation that diligent search be performed and recorded but, without clear legal guidelines, risk assessment had a key role to play¹. A solely risk-assessed approach to the digitisation of orphan works is still preferred by some organisations, while others use it in combination with the orphan works regime. In this section we consider some of the ways in which CHIs avoid, accept, mitigate and manage the risks associated with digitising copyright-protected material and making it available online. We consider risk assessment as an alternative to strict compliance with copyright law, and outline how we engaged in the process of risk assessing the Edwin Morgan Scrapbooks for this project.

2. COPYRIGHT AND DIGITISATION

The opportunities presented by widespread digital access to our shared cultural heritage are transformative. Recent surveys of the archive and library sector have found that users expect all library and archive materials to be digitised and available.² Of course, professional codes of conduct also state that archivists and librarians must respect the rights of individuals: they are committed to complying with the requirements of the Copyright Designs and Patents Act 1988, among other legislation. But, because of the restrictions of the legislation and the costs of rights clearance,³ it is

¹ See, for example, the module on Orphan Works and Risk Management available at: www.web2rights.com/SCAIPRModule/rlo3.html (accessed 10 January 2017), or in Pedley, P., *Copyright Compliance: Practical steps to stay within the law* (London: Facet Publishing, 2008).

² For example, see Dooley, Jackie M., Rachel Beckett, Alison Cullingford, Katie Sambrook, Chris Shepard, and Sue Worrall (2013), *Survey of Special Collections and Archives in the United Kingdom and Ireland*. Dublin, Ohio: OCLC Research. Available at: www.oclc.org/resources/research/publications/library/2013/2013-01.pdf (accessed 16 November 2016).

³ Research has shown that the cost of rights clearance usually outstrips both the cost of digitisation and the monetary value of the work itself: Vuopala, A., "Assessment of the Orphan works issue and Costs for Rights Clearance" (May 2010), available at: www.ace-film.eu/wp-content/uploads/2010/09/Copyright_anna_report-1.pdf (accessed 16 November 2016), p. 5; 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: www.jisc.ac.uk/publications/reports/2009/infromthecold.aspx (accessed 16 November 2016), p.21; Stratton, B., *Seeking New Landscapes: A rights clearance study in the context of mass digitisation of 140 books published between 1870 and 2010* (2011) London: British

clear that copyright plays a significant role in determining what material is selected for digitisation. Archivists either tend to select material that is in the public domain, or material for which they can be reasonably certain that their organisation owns the rights (or has been assigned them by a depositor with the authority to do so).⁴ One consequence of this copyright-driven selection process is illustrated,⁵ in part, by the 20th century black hole in the Europeana data set pictured below (Table 10): note the significant drop in the amount of material available through Europeana that was created from the 1950s onwards, as this material is more likely to be protected by copyright than earlier material.

Chronological distribution of *dcterms:issued* values in Europeana dataset (1800-today)

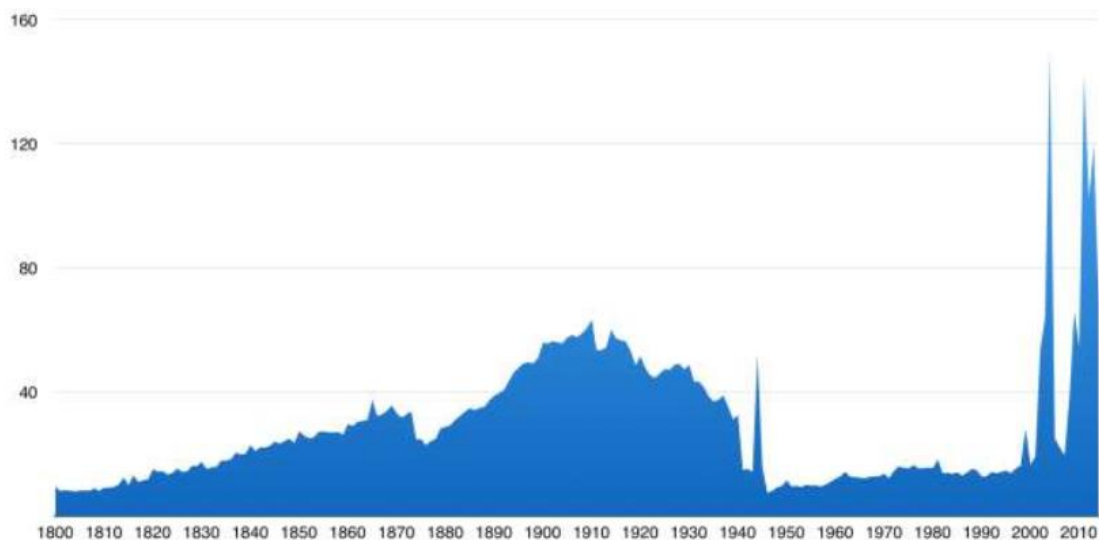


Table 10: Date range of materials in the Europeana dataset ⁶

There are two problematic issues with using rights status as a method to select works for digitisation. First, the proportion of public domain material in UK collections is significantly reduced compared to most other jurisdictions because of the 2039 rule.⁷ Second, asking depositors or donors to assign rights to institutions is not always possible: the depositor may not be the rightholder, and collections often contain third-party materials. Certainly, asking depositors to assign copyright where they can is advisable but staff may find that the agreements for older collections do not include

Library/ARROW. Available at: www.arrow-net.eu/sites/default/files/Seeking%20New%20Landscapes.pdf (accessed 16 November 2016), p.5.

⁴ Dryden, "Copyright issues in the selection of archival material for internet access" (2008) *Archival Science*, p.123.

⁵ Other factors will influence selection processes: staff skills and equipment, budget constraints, and designing the project to meet the needs of a funder, parent institution, or specific group of users.

⁶ Europeana Factsheet (2015) The 20th Century Black Hole: How does this show up on Europeana? Available at pro.europeana.eu/files/Europeana_Professional/Advocacy/Twentieth%20Century%20Black%20Hole/copy-of-europeana-policy-illustrating-the-20th-century-black-hole-in-the-europeana-dataset.pdf (accessed 17 November 2016).

⁷ For more information, see Legal Landscape Section 3, and www.create.ac.uk/blog/2014/06/02/will-uk-unpublished-works-finally-make-their-public-domain-debut/ (accessed 16 November 2016).

relevant assignments, or that documentation for older collections simply does not exist. As a result, many institutions choose to digitise material where they can be certain that they hold the copyright in the works selected, rather than digitising the material that best fits a particular research theme, user request or strategic priority for the institution.

Ultimately, the result of digitising material based on rights status is that users who access collections online are only ever seeing material that has been filtered through this selection process, rather than material which has been deliberately chosen to illustrate the full breadth and depth of the institution's complete holdings, and by extension our shared cultural heritage: from the oldest manuscripts through to born-digital records. In short, the digital historical record becomes skewed towards material that presents no or minimal rights clearance issues. This is a concern for at least three reasons. First, if digital is now the principal method of access to records for many CHI users, those users may not be aware of or attentive to the records that are absent from the digital collection. Second, the skew towards older public domain works means these tend to be the materials that shape research opportunities and activity; this is a particular problem in disciplines such as the digital humanities, which rely on large datasets to conduct research and where researchers are not always able to travel to relevant institutions in person. Third, it creates a fundamental barrier to digital access and therefore to the digital preservation of more recently-created works.⁸

3. RIGHTS MANAGEMENT IS RISK MANAGEMENT

Risk is typically expressed as the severity of an outcome (or the extent of a benefit resulting from an outcome) occurring, multiplied by the likelihood or probability of it occurring.⁹ Risk normally occurs as the result of interaction with uncertainty, for example: we may be uncertain about the rights status of material in our collections; about the likelihood of a rightholder making a complaint about the use of material; and about the likelihood of consequences, such as financial obligations or reputational damage, arising from a complaint. We may be willing to tolerate the risk of making material available despite uncertainty, on the basis that the benefits realised by digitisation outweigh the potential severity of any negative outcomes.

This formulation can often be difficult to apply to the outcomes of CHIs digitising copyright-protected collections as clear data on the rights clearance efforts from previous digitisation projects are not widely available, and very few case studies have been published. Additionally, there is no case law where UK CHIs have been sued for copyright infringement; allied with a lack of data on near-misses and complaints,¹⁰ this

⁸ Digital preservation is typically carried out when work is in such poor condition that it is close to being lost or suffering further permanent damage. But, digital preservation is also often triggered by a user request to make use of the work in some way, whether in the searchroom, for display in an exhibition, or as part of a research project. That is, delivering digital access is often the prompt for preservation activity. In this way, if copyright status presents a barrier to digital access, similarly it can also impede preservation activity.

⁹ For example, the Institute of Risk Management defines risk as "the combination of the probability of an event and its consequences. In all types of undertaking, there is the potential for events and consequences that constitute opportunities for benefit (upside) or threats to success (downside)." See Institute for Risk Management (2002) *A Risk Management Standard*, p.1, available at: www.theirm.org/media/886059/ARMS_2002_IRM.pdf (accessed 22 November 2016).

¹⁰ The authors define 'near-miss' in this context as a complaint about copyright infringement which could result in litigation, or where litigation is threatened, but which is resolved, either by negotiation or by capitulation, before proceedings are issued, or where proceedings are abandoned.

makes it difficult to predict the probability of litigation against a CHI or reputational damage occurring as the result of a complaint. That said, while the lack of litigation is a revealing metric in itself, in that it underlines the seeming unlikelihood of litigation arising within the heritage sector, we should be cautious of reading too much into this given the fact that reliable data on near-misses and complaints is unavailable.¹¹

Furthermore, the CHI sector could be more vocal and proactive in articulating the impact and value of digitised collections, which would make it easier to calculate the benefits of digitisation as against the risk of infringement. One way of doing this would be to use the Balanced Value Impact Model to articulate the different kinds of values, benefits and impacts generated by digitisation.¹² For example, by clearly articulating the social value of digitising local film collections, a local history museum could balance the benefits (improved user experiences, new outreach activities, an increased sense of place and belonging for participants, donations of film materials, increased knowledge about collections, and so on) against the risks (copyright infringement, sensitivity, complaints from rights holders, potential loss of good reputation). By doing this, they could then put in place strategies to minimise those risks and maximise the benefits; for example: by creating a local film history group; publicising the search for rights holders and the people who appear in the films; run screenings where viewers can provide feedback, information and memories; work with local social care providers to run memory sessions; and, contribute to local schools' learning resources.

Risk	Probability (1-5)	Severity (1-5)	Score (PxS)	Action to Prevent or Manage Risk
Legal	3	5	15	The EThOS and EThOSnet projects are addressing the legal aspects of collecting, digitising and making this type of material available.
More theses to be digitised than expected	5 (realised)	2	10	The project has delivered 4 times the number of theses to be digitised than originally expected. This means greater logistical involvement for the British Library, but the additional resource can be made available.
Institutions attempting to clear rights with authors	3 (small number realised)	3	9	A small number of institutions are contacting authors for clearance to make their theses available. 3 or 4 of the bigger institutions are doing this impacting on the logistics of the project. This is containable by applying a time limit of late May for decisions and the addition of further theses to replace those withdrawn.

Table 11: An example of a completed balanced scorecard for risk assessment.

¹¹ For example, the authors know of at least one action initiated against a UK archive institution, which was dropped before reaching court.

¹² Tanner, S. (2012) Balanced Value Impact Model.

There are three common ways of expressing and calculating these risks. One example of a traditional method is the balanced scorecard approach, which can be used as part of the project management process. A completed example, taken from a JISC digitisation project, is given above (Table 11).¹³ Users of this method are expected to assign a numerical value against the probability of an event occurring, with '1' meaning no to low probability and '5' meaning the event is highly likely to take place. A numerical value is then assigned to the severity of an event, with '1' meaning little to no effect, and '5' meaning severe consequences for project outcomes. The values for probability and severity are then multiplied to give a total score, and a section of the table is provided to record in detail how the risk identified will be mitigated or avoided. This allows project managers to see at a glance the project elements which carry the most risk and how they are being managed. This scorecard uses 5x5 scoring, but 3x3 and 'High, Medium and Low' scoring is also common.

The second approach adopts a similar methodology in that the scorecard was developed into a risk calculator, created by the Web2Rights project for an Open Educational Resources Toolkit. The risk calculator assigns numerical values to different types of material and the different ways in which they can be used, giving a high, medium or low ranking for different uses in addition to a numeric score. An example taken from the risk calculator, using an artistic photograph as the subject, can be seen below (Image 14).¹⁴

Artistic photograph

- Artistic photograph

Created with commercial intent?
Unknown

Includes clinical content/images of identifiable individuals/children for which rights have not been cleared?
No

Licence intended for your end product?
CC BY NC ND

Nature of creator
Known

Contact with Rights holder?
Known(low profile & traceable but don't respond)

Indicative Risk Level: 384

Medium risk = 151 – 500

This Indicative Risk Value indicates that the level of risk associated with the content that you wish to use is likely to require serious consideration. If your organisation is not comfortable with this level of risk, you should consider lowering the potential risk by providing access under a more restrictive end use licence and/or selecting material which reduces the Indicative Risk.

Image 14: Example of an artistic photograph and use ranked using the Web2Rights OER Risk Management Calculator

¹³ This scorecard is taken from the UK Thesis Digitisation Project Project Plan, available at: webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/media/documents/programes/digitisation/ukthesespp.pdf (accessed 17 November 2016) p.6.

¹⁴ The Risk Management Calculator is available at: www.web2rights.com/OERIPRSupport/risk-management-calculator/ (accessed 22 November 2016).

As we can see, the user of the calculator has decided to explore the risk associated with making an artistic photograph available. It is not known whether the photograph was created with commercial intent, but we do know that the photograph does not include clinical content, or images of identifiable individuals or children. The user wants to make the image available under a Creative Commons Attribution NonCommercial NoDerivatives licence. The creator is known, with a low profile, and the user has found contact details and approached the rightholder for permission, but they have not responded. The calculator gives a score of 384, which places it within the 'Medium' band (which includes scores of 151-500).

A third and final option is to define bespoke 'criteria' or 'categories' of risk for specific institutional digitisation projects; the Wellcome Library case study outlined below provides just such an example.

Whichever approach is adopted, it's important to bear in mind that rights clearance is cyclical and costs will build up throughout a project; that is, costs are not based purely on the amount of money that is paid to rightholders to obtain licences. Indeed, the vast majority of rightholders contacted during rights clearance for archive and library digitisation projects grant permission without requesting fees. Rather, the factors that can contribute to the potential costs range from the time taken to create item-level metadata at the beginning of a project, to maintaining records once the project is completed, and so on. For example, staff must: identify the copyright status of the works selected for digitisation, bearing in mind that one item can include multiple rights and therefore multiple rightholders; identify the rightholder(s) in the works (which could include estates, multiple heirs or successor companies); locate and contact the rightholder(s); negotiate permission for specific uses based on the way in which they intend the work to be use (this may or may not incur licence fees); and, create and maintain associated metadata and records of all of these actions, transactions and results.

4. CASE STUDIES: TAKING AND AVOIDING RISKS

In this section, we consider three published case studies concerning institutions that have taken different approaches to managing the risks associated with rights clearance. The first explores the Jon Cohen AIDS research collection at University of Michigan.¹⁵ The second is taken from the Southern Historical Collection and Carolina Digital Library and Archives, based at the University of North Carolina at Chapel Hill, where the Thomas E Watson Papers were digitised.¹⁶ The final example is the Wellcome Library (working with five partner institutions) and its pilot mass digitisation project, Codebreakers: Makers of Modern Genetics.¹⁷ The commentary within this section is arranged thematically comparing and contrasting the actions of each institution at each stage in the digitisation process.

¹⁵ Akmon, D. (2010), "Only with your permission: how rights holders respond (or don't respond) to requests to display archival materials online", *Archival Science*, 10(1), pp. 45-64.

¹⁶ Dickson, M. (2010), "Due Diligence, Futile Effort", *The American Archivist*, 73(2), pp. 626-636.

¹⁷ Stobo, V., Deazley, R., and Anderson, I.G. (2013) "Copyright and Risk: Scoping the Wellcome Digital Library", Working Paper 2013/10, CREATE, University of Glasgow, Glasgow, available at: zenodo.org/record/8380/files/CREATE-Working-Paper-2013-10.pdf (accessed 25 November 2016).

4.1. COLLECTIONS BACKGROUND

The collections include: the Jon Cohen AIDS research collection, comprising the personal papers of Jon Cohen, a journalist and science writer who covered the development of a vaccine for the condition during the 1980s; the personal papers of Thomas E Watson, a prominent US senator active in the late 19th and early 20th century; and 20 collections of personal papers created by geneticists of the 20th century, including James Watson, Francis Crick, and Rosalind Franklin, among others.

4.2. RIGHTS AUDITING THE COLLECTIONS

Staff at the University of Michigan audited the Jon Cohen collection and found 13,381 items. 6,026 items (45%) were found to be either newspaper or journal articles, and the library staff decided not to digitise these as the majority were already available online elsewhere. 1,892 (14%) of the items were US Federal Government documents¹⁸ which could be made available without seeking permission. For only 209 documents was Jon Cohen the copyright holder (2%), for which he granted permission. This left the staff with the task of securing permissions for the remaining 5,254 items (39%) where copyright was held by a third-party. The staff identified 1,376 unique rightholders in this material.

The archivists at University of North Carolina sought and were granted permission to digitise the material created by Thomas E. Watson from his surviving heirs. They audited the collection and found 3,304 correspondents, with identifying information for 3,280 correspondents. They found birth and death dates for 1,709, leaving 1,571 with no dates. 1,101 of the correspondents died after 1939, which was the cut-off date the archivists had decided to use in determining whether a letter was still protected by copyright or not. 608 died before 1939, allowing the archivists to assume that these letters were in the public domain.¹⁹ This process took four and a half months and cost almost \$6000. At the end of this process, the archivists determined that the collection could be categorised as follows: 14% were created by Watson family members and permission had been granted by the estate; 3% were created by freelance workers and excluded from the scope of the project because of time and expense; 4% of the letters were unsigned, illegible or used pseudonyms; 21% of the letters were in the public domain; 27% were still in copyright; and 31% of the letters had an undefined copyright status.

The Wellcome Library staff and their partners realised there would be many thousands of third-party rightholders across the 20 collections of personal papers selected for digitisation: altogether, almost 3M pages were digitised. To manage this process, they asked their partner archives to identify rightholders in their collections according to a set of risk criteria shown in the table below (Table 12). Using these criteria, the Wellcome and their partners were able to iteratively reduce the number of rightholders they intended to contact from thousands to just 160 across all the collections.

¹⁸ US Federal government documents do not benefit from domestic copyright protection, and are therefore in the public domain (s.105 of the 1976 Copyright Act).

¹⁹ This digitisation project took place in 2009. The assumption that material created pre-1939 would be in the public domain is based on the US copyright term of life plus 70 years: 1939 + 70 = 2009.

HIGH RISK	MEDIUM RISK
Author is a well-known literary figure, broadcaster, artist	Author has (or had) a high public profile
The author/estate/publisher is known to actively defend their copyright	Author is alive and known to have a literary estate as recorded in the WATCH file
The relationship between the institution and the author/estate/publisher is awkward	The material appears to have been published or broadcast and/or prepared for commercial gain rather than to advance academic knowledge or in a not-for-profit context

Table 12: Risk criteria used by the Wellcome Library during Codebreakers: Makers of Modern Genetics

4.3. THE RIGHTS CLEARANCE PROCESS

Where contact details were found, the staff at the University of Michigan sent the rightholder a letter that explained the project, described the material they wanted to digitise, and asked for a non-exclusive licence to digitise the item and make it available online, without the offer of a licence fee. The letter also included a statement of support from Jon Cohen and the funders for the project, along with a consent form.

The staff at the University of North Carolina attempted to find identifying information for rightholders using 'ancestry.com, the Congressional Biographical Directory, the Historical Marker Database online, the Library of Congress authority database, the New Georgia Encyclopedia, print references, the Social Security Death Index, the WATCH File, Wikipedia, and WWI draft registration forms.'²⁰

Wellcome Library staff searched Who's Who, the WATCH File, Google, Wellcome Trust internal databases, third-party archives, the Dictionary of National Biography, obituaries and Wikipedia to find contact details for rightholders and their heirs.²¹ The results of this process, including initial contacts, follow-up attempts and details of permissions granted and refused were carefully documented. Rightholders were sent a letter that explained the project, outlined the material to be digitised, and asked for permission to make the material available under a non-exclusive Creative Commons Attribution Non-Commercial 4.0 licence. The Wellcome did not offer a licence fee to rightholders.

Staff from both Michigan and the Wellcome sent follow-up letters and used email and phone where possible to secure permissions. Many rightholders requested copies of the material before making a decision, and these were sent by email, fax or post where required. It is worth noting that such requests are common during digitisation projects and time should be factored in at the project planning stage to accommodate them.

²⁰ *ibid*, p.629. The article does not provide information on the content of the permission letters.

²¹ Stobo, V. et al (2013) p.26.

4.4. RESULTS OF THE RIGHTS CLEARANCE PROCESS

The following table presents the results of the rights clearance process at each of the case study institutions (Table 13).

	Jon Cohen AIDS Collection	Thomas E Watson Papers	Codebreakers
No. of Copyright Owners identified	1,377	3,280	160
No. Copyright Owners traced	1,023 (74% of those identified)	4 (0.12% of those identified)	134 (84% of those identified)
Replied	748 (68% of those contacted) ²²	3 (75% of those contacted)	103 (77% of those contacted)
Permission granted for all items	679 (91% of respondents)	3 (100% of respondents)	101 (98% of respondents)
Permission granted for some items	23 (3% of respondents)	n/a	n/a
Permission denied	46 (6% of respondents)	n/a	2 (2% of respondents)
Non Response	352 (32% of those contacted)	1 (25% of those contacted)	26 (19% of those contacted)
Orphan Works	354 (26% of those identified)	3,276 (99.88% of those identified)	22 (14% of those identified)
In Progress	n/a	n/a	4 (4% of those identified)

Table 13: Rights Clearance Results

We can see that the archivists at the University of Michigan had a success rate of 74% in finding contact details for rightholders. 68% of those contacted responded, with 94% of respondents granting permission for all or some of the material requested. 6% of respondents refused. 32% of those contacted did not respond, and contact details could not be found for 26% of the rightholders identified in the collection. The archivists decided not to make non-respondent or orphan material available online. If we add these items to those where permission was refused, the University of Michigan was unable to make 1,973 items available, or 36% of the total collection. This is despite receiving dedicated funding for the digitisation and rights clearance process, and spending, on average, 70 minutes per rightholder on securing express permissions. Moreover, the study revealed that commercial rightholders are more likely to refuse

²² The figures for responses do not tally completely because 83 rightholders were contacted on two occasions and were asked for permission to digitise two different sets of items: those permissions were recorded separately.

permission requests, and that after four months, subsequent permission requests deliver diminishing returns.

The archivists at the University of North Carolina managed to find contact details for the estates of only four correspondents out of the 3,280 identified, either by using the WATCH file or by contacting other repositories where manuscript collections were known to contain material created by identified correspondents. Of the four, three estates granted permission and one did not respond.

Staff at the Wellcome managed to find contact details for 84% of the selected rightholders. 77% of those contacted replied, with 98% of respondents granting permission. 19% of those contacted did not respond, but after re-assessing the likelihood of those rightholders objecting to publication, most of the material was made available online, subject to takedown requests. Contact details could not be found for 14% of the selected rightholders, meaning that this material is orphaned. In contrast to the University of Michigan, the Wellcome decided to make orphan work and non-respondent material available online in batches, excluding material which was deemed to be very high risk. As a consequence, staff were able to make most of the collections they selected for digitisation available online: indeed, the material associated with 91% of the selected rightholders was made available, in addition to all of the other material where rightholders were not contacted at all.

4.5. RISK MANAGEMENT DECISIONS

The University of Michigan decided to adhere to a policy of strict copyright compliance and only make third-party material available where they had been granted express permission. As a result, they made no orphan works or non-respondent material available. At the end of the project, almost 36% of the collection was not available to view online.

If the archivists at the University of North Carolina had followed the same path, they would only have been able to make available online the Watson family material, public domain material, and material for which they were able to secure permission: just 35% of the total collection. However, rather than digitise only when express permission was granted, the archivists turned to the US copyright doctrine of fair use.²³ Following the initial attempts to clear rights in the material, the archivists involved in the project

²³ Section 107 of the US Copyright Act 1976 provides that ‘the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright’. In making a determination about whether an act constitutes fair use In determining whether the use made of a work in any particular case is a fair use, s.107 requires that the following four factors be taken into account: ‘the purpose and character of the use, including whether such use is of commercial nature or is for non-profit educational purposes; the nature of the work itself (whether it is a factual or creative work); the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and the effect of the use upon the potential market for or value of the copyrighted work’.

Fair Use justifications like those used by University of North Carolina have been used to facilitate many digitisation projects at US CHIs. Such justifications still rely on elements of risk management, given the need to interpret the four factors. See Aufderheide, P. and Jaszi, P. (2011) *Reclaiming Fair Use: How to put balance back in Copyright*, Chicago: University of Chicago Press for more details, and the *Code of Best Practices in Fair Use for Academic and Research Libraries* (available at: www.arl.org/storage/documents/publications/code-of-best-practices-fair-use.pdf) for an example of fair use guidance.

presented their findings to the legal counsel for University of North Carolina University Libraries and explained that they wanted to discontinue any further copyright investigation for this collection. In turn, they were given permission by the University to make the material available online: the University had been convinced by the fair use argument.²⁴ A year later, staff reported that no takedown requests had been received in relation to the Thomas Watson material.

Given the huge size of the combined collections selected for digitisation, the Wellcome Library staff and their partners realised that contacting all third-party rightholders would be a significant undertaking with little guarantee of comprehensive success. Unlike the staff at the University of North Carolina, they could not avail of a generalist defence like the US fair use argument; they understood that, for the project to be successful, they would have to accept a much larger degree of risk. They managed this by developing the criteria for identifying rightholders likely to object to publication, which they combined with a 'best endeavours' search to trace and contact those rightholders. The Wellcome Library also has a takedown policy that applies to all of the material made available on their site. To date, the Wellcome Library has received only two takedown requests in relation to Codebreakers material. No reasons were given for the requests, no compensation was sought by the requestors, and no litigation has ensued.²⁵

4.6. ANALYSIS

The staff involved in each of these rights clearance exercises made decisions to set strategic boundaries for their respective projects. For example, the staff at both the Wellcome and the University of Michigan decided not to digitise newspapers and journals for two reasons: it was considered to be a waste of effort where the material was available elsewhere, and the rights clearance process for the material was perceived to be too onerous. The University of North Carolina decided to exclude material made on a freelance basis from their project, on the grounds that including the material would have created further time and expense for the archivists involved in rights clearance.²⁶ At the University of Michigan, non-response was taken as denial of permission, while the University of North Carolina decided to make non-responders' materials available on the basis that digitisation amounted to fair use. In contrast, the Wellcome decided to make most of their non-respondent and orphan material available, unless it was deemed to be very high-risk.

CHIs must be aware of the trade-offs in making such decisions, balancing resources and perceived risks against the benefits of making more collections available online. As previously mentioned, digitisation efforts should be focused on the most appropriate material for a particular project. In turn, staff at cultural institutions should consider weighing the risks of making copyright-protected material available online without express permission against clearly-articulated benefits of doing so; with this in mind,

²⁴ Dickson, 636.

²⁵ Stobo et al (2013) Copyright and Risk: Scoping the Wellcome Digital Library Project, CREATE Working Paper 2013/10, CREATE, University of Glasgow, available at: (accessed 5 December 2016).

²⁶ Including material made on a freelance basis would have generated more work for the project archivists, as they would have had to assess whether the copyright was held by the freelance worker, or by the employer the freelancer was working for. Usually copyright vests with the freelance worker in such situations, but this can be subject to an agreement to the contrary, where the copyright will be retained by the employer. Assessing all the freelance works in the collection would have taken up considerable time with no guarantee of successful clearance.

they should formulate strategies that minimise the risks and maximise the benefits. The case studies discussed in this section illustrate the potential value of risk-based strategies, strategies that we believe will become increasingly significant as more institutions digitise their collections in the absence of meaningful legal reform in this arena.

5. MANAGING RISK IN THE SCRAPBOOKS

Created in the mid-20th century from predominantly published material, the Edwin Morgan Scrapbooks contain a huge number of works that are still in copyright. Taking a 30-page sample (10%) of Scrapbook 12, we carried out a full data extraction exercise that included assessing the risk level of each item. From this sample of 432 works, 64% were in copyright to both known and unknown rightholders.²⁷

Throughout the assessment process, uncertainty was a constant undertone. Aside from the concerns of carrying out a search that would be sufficiently diligent, the incomplete nature of many of the works presented a challenge. From the sample, 50% of works were either incomplete or it wasn't possible to tell if they were a complete work. Of these, 35% were clearly incomplete, whereas for 15% it was not possible to determine definitively whether they were incomplete or not. With this in mind, the 'completeness' of each cutting was assessed as: Yes-No-Unknown.

Some of the works that were incomplete could be deemed to be an insubstantial part of the larger work, allowing for lawful use without permission. However, incompleteness and insubstantiality are not the same thing: a very small cutting might display the most significant or recognisable part of an artistic or literary work even though it is quantitatively insignificant. For example, the extract might contain the most famous passage from a novel; as such, it would almost certainly be regarded as a substantial part of the work itself. In some cases, however, it was relatively easy to determine that a cutting was not substantial, for example, if it contained a few words of non-specific text or an indistinct section of an image.

Naturally, substantiality needs to be decided on a case-by-case basis, with consideration given to the content of the work and its context; this can be a daunting task, with many works falling within a grey area. From our sample of 432 works, we categorised 217 as either incomplete or unknown; of these, 84 (19% of sample total) were deemed to be insubstantial in nature.

Thereafter, we addressed the risk of making these materials available online without permission, adopting three risk categories: low, medium and high.²⁸ As with the issue of substantiality, making this determination was not always clear-cut. We discuss our risk assessment categories in the next section.

Images 15, 16 and 17 illustrate some of the challenges posed by the material contained in the Scrapbooks. All three were considered by the Project Officer to pose a low risk from the perspective of digitising for use online. Readers may or may not agree. The first

²⁷ For these purposes we exclude works created by Morgan himself, works in the public domain, insubstantial parts of works that may or may not be in copyright, and ephemera. For further details, see *Diligent Search in Context and Practice*, Table 9.

²⁸ Sometimes, items classed as low risk could also be described as 'no risk,' as this category includes works which are not subject to copyright protection, as well as works for which rights were subsequently cleared as part of the digitisation project.

cutting (Image 15) is small in size,²⁹ taking up approximately 1/16 of the scrapbook page. The image might be a part of a larger image. For example, the right side of the cutting looks misshapen, as if it had been cropped from a larger picture. But, this is speculation only. All that can be said is that the image composition does not give any meaningful indication as to whether the image was originally intended to be as shown. The image may or may not be complete, or it may



Image 16: Colour image of waves (perhaps)

be a substantial part of a larger work. It is not possible to tell.

The second example is a colour image of what appears to be waves (Image 16). The image is small, taking up less than 1/16 of the page. It is definitely cropped from a larger image. The subject matter is indistinct and difficult to make out; as such, we considered it to be insubstantial.



Image 15: Colour photograph of a man operating a control panel

Together they take up less than 1/16 of the scrapbook page. Both are definitely incomplete, and it is likely that the television surround came from a magazine advert (such things are used elsewhere in the scrapbooks). By and large, cuttings from magazine adverts of this kind were considered to be ephemera and extremely low risk. The black and white photograph appears to have been greatly cropped. As such, we considered it likely to be an insubstantial part of the larger work. As with the other two cuttings, we determined that the risk of making use of the image without permission was low.

The third (Image 17) consists of two images. One is the wooden frame of the television and the other is the black and white image of the figure inside.



Image 17: A collage

5.1. RISK ASSESSMENT CATEGORIES

After an initial assessment of the scrapbooks by the Project Officer, a set of risk categories were developed to allow categorisation of the material during the data extraction process on the 10% sample. The criteria were developed partly in response to the type of material found in the scrapbooks but also taking into account more general risk criteria informed, in part, by the Wellcome Codebreakers digitisation project.³⁰ In

²⁹ See, for example, the newspaper text above the image which provides a sense of scale.

³⁰ For details, V. Stobo with R. Deazley and I.G. Anderson, *Copyright & Risk: Scoping the Wellcome Digital Library Project* (2013) CREATE Working Paper 2013/10, DOI: 10.5281/zenodo.8380.

Table 14 we set out the guidelines underpinning the risk classification of each item from the sample.

NO/LOW RISK	<p>Item was created by Edwin Morgan</p> <p>Item is no longer in copyright</p> <p>Item is a piece of ephemera ³¹</p> <p>Item appears to be an insubstantial part of a larger work</p> <p>Published works authored by private individuals for a non-commercial purpose, e.g., a letter written to a newspaper or magazine</p>
MEDIUM	<p>Personal photographs or other similar items, not produced for commercial purposes, where the author is thought to be a friend or correspondent of Morgan</p> <p>Work thought to be in copyright where an author is named but no further information can be found</p> <p>Work thought to be in copyright with an identifiable publisher that no longer appears to exist</p>
HIGH	<p>Item is still in copyright and the rightholder is identifiable</p> <p>Author is known to have an active estate/publisher defending copyright</p> <p>Substantial extract from a book or article, and particularly when other extracts taken from the same work appear elsewhere in the Scrapbooks</p>

Table 14: Risk Guidelines from Edwin Morgan Project

During the data extraction process, each item was initially classified as either low, medium or high risk. Following this initial classification, the Project Officer targeted the high risk material to see if permission could be secured for use. As we noted in *Diligent Search in Context and Practice*, a number of rightholders (15) granted permission to make use of their work free of charge; whenever permission was secured the relevant work was re-designated ‘no risk’. In addition, five orphan works were cleared for use through OWLS; they too can be considered ‘no risk’, at least until the seven-year licence expires. A further five orphan works were cleared for use through the Directive.

After more than 1000 hours spent on diligent search and rights clearance activity by the Project Officer, 61% of the sample was deemed to be no/low risk, 34% was medium risk, with 5% remaining in the high risk category. Naturally, the high risk category includes each of the 19 works for which the relevant rightholder was only prepared to grant permission conditional on payment of a fee. No licence fees were paid.

³¹ Ephemera are items without lasting significance, intended to be used for a short period of time e.g. tickets or advertisements. For this reason they can be classed as low risk, as long as they do not contain material from a source still in evident copyright, such as a photograph or artwork. Even if these ephemera are protected by copyright, the owner is highly unlikely to object to use (so long after the event) given that these works were originally created for functional/informational purposes at the time

6. CONCLUSION

Rather than present conclusions within each section, we collect them together within one project conclusion (available at <http://www.digitisingmorgan.org/conclusion>).

This document is a section taken from the website www.digitisingmorgan.org and will also be included in a forthcoming CREATE Working Paper. The site was created as part of *Digitising the Edwin Morgan Scrapbooks*, through support by the RCUK funded Centre for Copyright and New Business Models in the Creative Economy (CREATE), AHRC Grant Number AH/K000179/1.

Date Created: January 2017

Cite as: V. Stobo, K. Patterson and R. Deazley, *Digitisation and Risk* (2017), available at www.digitisingmorgan.org/risk