DIGITISING THE EDWIN MORGAN SCRAPBOOKS:
DILIGENT SEARCH IN CONTEXT AND PRACTICE

Kerry Patterson, Ronan Deazley and Victoria Stobo

1. INTRODUCTION

Carrying out a diligent search is a requirement of both the EU Orphan Works Directive[1] and the UK Intellectual Property Office's Orphan Works Licensing Scheme (OWLS). Diligent search is a time-consuming exercise for any digitisation project and the task is even more challenging with the Edwin Morgan scrapbooks due to the number of de-contextualised and partial cuttings and the variety of sources used. In this part of the resource we explore the practical implications of diligent search in relation to both the Directive and OWLS, illustrated by a rights clearance exercise performed on a 10% sample of Scrapbook 12 from Edwin Morgan's series of scrapbooks.

2. DILIGENT SEARCH IN CONTEXT

This project is the first major UK study concerning the concept of diligent search since the Directive and OWLS came into effect. The costs and challenges of rights clearance activity and of dealing with orphan works have been identified as significant barriers to the digitisation of cultural heritage collections by various studies.[2] We begin by discussing a selection of those studies.

Denise Troll Covey's 2005 research reports the results of Carnegie Mellon University Library's 'feasibility study to determine the likelihood of publishers granting nonexclusive permission to digitize and provide surface Web access to their copyrighted books.'[3] An initial sample of 368 books selected at random was reduced to 277 titles; the reduction was attributable to mis-cataloguing and other reasons, including that a substantial number of works were deemed too complicated to include in the study because of third-party copyright issues (11% of the initial sample).[4] From the sample of 277 items from 209 publishers, a rightholder could not be located for 19% of the books.

---

[1] DIRECTIVE 2012/28/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 25 October 2012 on certain permitted uses of orphan works; see http://ec.europa.eu/internal_market/copyright/orphan_works/index_en.htm#maincontentSec1 (accessed 13 December 2016)

[2] See for example: Deazley, R & Stobo V (2013) *Archives and Copyright: Risk and Reform*, available at https://zenodo.org/record/8373/files/CREATe-Working-Paper-2013-03.pdf (accessed 22 November 2016); Korn, N (2009) *In from the Cold: An assessment of the scope of 'Orphan Works' and its impact on the delivery of services to the public,* available at https://sca.jiscinvolve.org/wp/files/2009/06/sca_colltrust_orphan_works_v1-final.pdf (accessed 22 November 2016); Vuopala, A (2010) *Assessment of the Orphan works issue and Costs for Rights Clearance,* available at http://www.ace-film.eu/wp-content/uploads/2010/09/Copyright_anna_report-1.pdf (accessed 22 November 2016); Akmon, D (2010) *Only with Your Permission: How Rights Holders Respond (or Don't Respond) to Requests to Display Archival Material Online,* available at http://link.springer.com/article/10.1007%2Fs10502-010-9116-z (accessed 14 December 2016)

[3] 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 13 December 2016), 11

[4] Ibid.

Of the remaining works, 27% of rightholders did not respond to the requests for permission to digitise while 30% expressly denied permission. Less than a quarter of rightholders (24%) eventually agreed to allow digitisation.

It is worth nothing the length of time and labour required to achieve these results. The average length of time to receive a response from a publisher ranged from an average 101 days for a response of 'Permission granted,' to an average 124 days for a response of 'Permission denied.' Moreover, more than 60% of publishers required a second or third letter before responding; in total, 524 letters were sent.[5] The labour and material costs of searching for rightholders were estimated at $200 USD per cleared work, not including agreed fees.[6]

Barbara Stratton's 2011 study focussed on rights clearance for 140 books published between 1870 and 2010 and held by the British Library.[7] Directly inspired by Covey's work, one of the report's stated goals was to 'measure and quantify the level of diligent search currently required to undertake mass digitisation of material from the last 140 years,'[8] in addition to identifying copyright status and the proportion of orphan works.

Before any rightholders were contacted, initial information had determined that of the total sample, 57% (80 books) of the books were in copyright, 27% (38 books) were in the public domain and the remaining 16% (22 books) had an unknown copyright status. Therefore rights clearance was investigated for the 73% of books (102 books) not believed to be in the public domain. These figures changed very little following rights research: the number of unknown works dropped by 2% when two books were confirmed to be in copyright and a further two works were confirmed as in the public domain.[9]

Following the period of rights clearance and rights research, it was determined that 43 of the books were orphan works, equating to 31% of the total sample of 140 books. Of these, 15% (21 books) were definitely still in copyright, while 16% (22 books) were orphan works of unknown copyright status.[10] On average, it took four hours to complete a diligent search for each title and more than 450 hours (80% of the project time) were spent on research and identification tasks before a single rightholder was even contacted.[11]

Taken together, Covey and Stratton's studies set out issues common to many digitisation projects, beginning with the 'weeding out' of the most complicated items – from a copyright perspective – at the beginning of the study, as described by Covey. Both studies evidence a high proportion of orphan work material within each of the random samples (31% by Stratton and 19% by Covey), a result that is particularly striking given

---

[5] Comprising 278 initial request letters and 246 follow-up letters; Covey, *Acquiring copyright permission to digitise and provide open access to books,* 13.

[6] These are noted to be transaction costs and do not include permissions paid to publishers for the rights to digitise for use online. Permission fees were paid up to $100 per item.

[7] Stratton, B. (2011), *Seeking New Landscapes: A rights clearance study in the context of mass digitisation of 140 books published between 1870 and 2010, Project Report for British Library/ARROW,* available at: http://www.arrow-net.eu/sites/default/files/Seeking%20New%20Landscapes.pdf (accessed 13 December 2016).

[8] Ibid., 4.

[9] Ibid., 33.

[10] Ibid., 37.

[11] Ibid., 51.

that both studies concerned published materials for which one might expect more reliable, locatable data concerning rightholders. Time and the other associated costs also are significant in each study, with four hours per item for rights clearance activity in Stratton's study in addition to the 80% of project time spent on pre-clearance research. Covey's study produced similar results with an average $200 cost per item. These figures provide little comfort for any institution contemplating a digitisation initiative of any scale.

The title of Maggie Dickson's case study of the digitisation of the Thomas E. Watson papers, 'Due Diligence, Futile Effort,' sums up the experience of digitising a collection of late nineteenth and early twentieth century manuscripts.[12] A pilot for a larger digitisation project, one of the principal lessons that emerged was that using the methods adopted in the pilot to clear rights on a larger scale would be 'needlessly expensive and futile.'[13] Rather, the study led Dickson to conclude that risk had a central role to play in future digitisation initiatives within the sector: 'If we are willing to calculate and assume some degree of risk and to document our decisions, archives and libraries can move forwards with large-scale digitization.'[14] Indeed, the issue of risk management is impossible to disentangle from any digitisation process, and can be a stumbling block for many institutions.

One further rights clearance study by Dharma Akmon highlights the familiar issue of the complexity of copyright law, but also mentions the lack of case law in this area[15], making it challenging for cultural heritage institutions to measure risk without examples from other organisations.[16] In the cited case to digitise the Jon Cohen AIDS Research collection, 85% of staff time on the digitisation project was spent on copyright permissions, with the average time of 1 hour 10 minutes per item,[17] figures very similar to those described by Covey and Stratton. Moreover, 'collections with a higher document to copyright holder ratio will probably cost less to usher through the rights process than collections with a low document to copyright holder ratio.'[18] This is certainly of relevance to the Edwin Morgan scrapbooks, where a single page could contain up to 30 individual rightholders.

As with Dickson, Akmon highlights that risks must be taken, emphasising that the unwillingness to take risks in the Cohen project meant that 18% of copyright items could not be displayed due to non-responding rightholders and a further 12% were not displayed due to unidentified rightholders[19]. We discuss risk in more detail in the next section of this resource. For now, suffice to say that even with the orphan works regimes introduced by the Directive and OWLS, managing risk remains a necessary aspect of any diligent search and digitisation activity.

---

[12] Dickson, M (2010) *Due Diligence, Futile Effort: Copyright and the Digitization of the Thomas E. Watson Papers,* available at http://americanarchivist.org/doi/10.17723/aarc.73.2.16rh811120280434 (accessed 22 November 2016).

[13] Ibid., 636.

[14] Ibid.

[15] Akmon, 2.

[16] Ibid., 2.

[17] Ibid., 26.

[18] Ibid., 26-27.

[19] Ibid., 27

3. DEFINING DILIGENT SEARCH

In *Orphan Works: The Legal Landscape* we provide a comprehensive analysis and critique of the orphan works regime, including the concept of diligent search, under both the Directive and OWLS. We will not revisit that commentary here, other than to reiterate that diligent search is fundamental to both the Directive and OWLS. Moreover, within the UK, the implementing legislation defines diligent search differently in relation to the orphan works exception (which requires a diligent search to be carried out 'in good faith') and OWLS (which requires the diligent search to be 'reasonable'). We interpret these standards to mean different things; that is, we consider the concept of a reasonable search conducted under OWLS to set a higher threshold than that of a good faith diligent search conducted under the Directive.

Importantly, the UK Intellectual Property Office (the IPO) has produced published guidelines on conducting diligent searches in relation to (i) film and sound-related orphan works, (ii) literary orphan works, and (iii) still visual art orphans.[20] These guidelines are 'primarily intended' for those wanting to make an application through OWLS, although each does state that they 'may also be of help to those conducting a diligent search in relation to the EU Directive'.[21] In addition, the guidelines are accompanied by 'Diligent Search checklists'; the checklists set out the key organisations to approach regarding possible orphan works. Taken together, the guidelines and the checklists provide very useful information and signposting when undertaking a diligent search.

Ultimately, however, the nature and demands of a diligent search will depend on the content and context of each project; there are no hard and fast rules that apply in every scenario. As the IPO makes clear in its guidance: there is 'no set way to conduct a diligent search as this will depend on the information available on the work,' and 'there is no minimum requirement to be followed in every case.'[22]

4. DATA EXTRACTION AND CATEGORISING WORKS

It was decided to carry out a rights clearance process for a 10% sample of Scrapbook 12, a 30-page long section, informed by the IPO's official guidelines on diligent search.

The first step was a process of data extraction, in which we recorded details for each of the 432 works present in the 380 individual cuttings in the sample.[23] The information recorded included type of material and its completeness (if possible to determine), as well as information on creator, source of publication, date and country of origin. Some of

---

[20] Under both the *Orphan Works Regulations* and the *OWLS Regulations* the Intellectual Property Office are empowered to produce on the appropriate sources to be consulted when conducting a diligent search. These guidelines are currently available here. The IPO is committed to reviewing and revising these guidelines as appropriate, and indeed they have already been updated since OWLS was launched.

[21] See, for example, Intellectual Property Office, *Orphan works diligent search guidance for applicants: Literary Works* (November 2015), 1, available here: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/474401/orphan-works-literary-works.pdf (accessed: 12 December 2016).

[22] The IPO provide a comprehensive list of resources, with guidelines for each category of orphan work – Film and Sound, Literary Works and Still Visual Art. Each document is 44 to 50 pages long and has been compiled by experts in the relevant sectors.

[23] Some items had multiple instances of copyright e.g. a cutting with text and a photograph or a photograph of an artwork in which the artwork and photographer both have rights.

this information was included within the cutting or had been provided by Edwin Morgan as an annotation, such as the date of a newspaper cutting. Other information could be determined based on other contextual information (for example, recognising the font of a magazine that was used and named elsewhere in the Scrapbooks) or was found following research by the Project Officer.

This process allowed the Project Officer to distinguish between possible orphan works and those works that did not require any permission to digitise and make available online. The items in the latter category could be described as material that was without, or was not likely to attract copyright protection: items in the public domain, ephemera and cuttings that were deemed to be an insubstantial part of the original work. For these items, no permission was needed to digitise and make the work available online, so it was not necessary to consider the orphan works regime. That said, it was not always easy to decide whether a given cutting is substantial or not without having access to the original work. For the rights clearance process, the Project Officer recorded 'completeness' of cuttings, a concept not completely synonymous with 'substantiality' but which was nevertheless helpful in making determinations about the need to rely on the orphan works regime or not. In the sample, 50% of items were complete, 35% were incomplete with 15% unknown in that it was not possible to make a definitive determination.[24]

5. PREPARING FOR DILIGENT SEARCH: CUTTINGS IN CONTEXT

Prior to starting a diligent search, an applicant should check the UK IPO's Orphan Works Register and the EUIPO's Database to ensure that the work in question is not already registered. As the schemes have only been in operation since 2014, the number of works registered with each is still relatively small: 1,950 works of all types on the EUIPO register and 456 on the OWLS database.[25]

As with the digitisation of many collections, and as advised by the OWLS Guidelines, looking to the provenance of the scrapbook gives an idea of the most appropriate sources to check. As Morgan compiled Scrapbook 12 between 1954 and 1960, we began by limiting our date selection for searching to roughly the decade between 1950 and 1960. Morgan does occasionally include older material, such as antique photographs or news cuttings (Images 1 and 2), however, these tend to be evident by their appearance; plus, Morgan often noted the year of non-contemporary material.

Image 1: Original antique photograph

Morgan was a voracious consumer of print media. He read English Language and Literature at Glasgow University, but was also familiar with French, Russian, Italian and German; items in these languages appear throughout the scrapbooks. His personal papers are housed in the University of Glasgow Library, however, they do not include a list (of any kind) of the sources Morgan used in creating his scrapbooks.

---

[24] For further discussion, see Digitisation and Risk, Section 5: Managing Risk in the Scrapbooks.

[25] OWLS database available at https://www.orphanworkslicensing.service.gov.uk/view-register (accessed 14 December 2016). Not all works registered on OWLS are granted a licence – some are pending and others are eventually withdrawn.
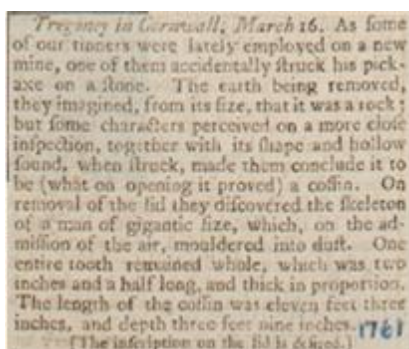
Image 2: Original newspaper cutting from 1761

That said, Morgan does provide some clues regarding some of the cuttings used. For longer length text cuttings, he wrote the initials of the newspaper and date on the cutting, or occasionally the full name of a published source (Image 3). Therefore, with knowledge of the types of sources Morgan favoured, gained through wider reading of the scrapbooks and using knowledge of newspapers that would have been available at the time, we were able to determine that 'GH' is *Glasgow Herald* and 'EN' is *Evening News*. Similarly, periodicals like *Life Magazine* or the *Illustrated London News* were often used by Morgan as sources for both images and stories.

The likely source of some uncredited text items can be found by looking at the typescript, the type of paper or even the subject matter, in order to match it up with known sources Morgan frequently used. For example, cuttings from *Doubt* magazine (Image 4), a favourite but niche periodical produced by the Fortean Society, can be identified by its distinctive font and glossy paper, combined with the subject matter of unexplained and paranormal events.
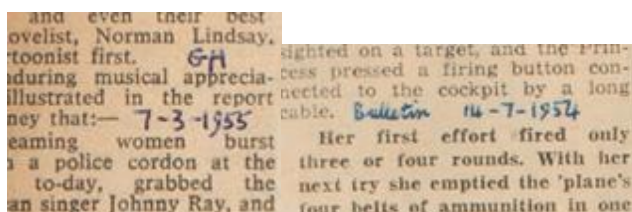

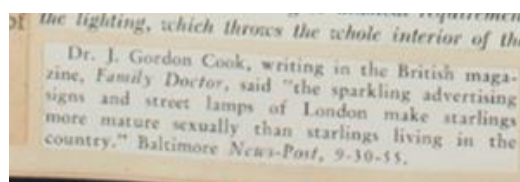**Image 3:** Cuttings from the Glasgow Herald and Bulletin, annotated by Morgan

Image 4: Cutting from Doubt Magazine

6. DILIGENT SEARCH: SEARCHING FOR TEXT ONLINE

Text-based searching using online search engines is a familiar and intuitive technique when trying to identify an unknown work, its creator or copyright owner.

For text-based cuttings included within the sample, Google yielded occasional results. Google News Archive has scans of some newspapers but this is not fully text searchable across all publications, although this capability is in development.[26] A list of all available newspapers available through the Google News Archive can be found here.[27] The News Archive digitisation project has ended, with no plans to add any more publications.

Magazines are also available through Google, as part of the Google Books section (for a complete list see here[28]) but, like the newspapers, only some are text searchable. For example, a search of text from *Life Magazine* found the source article, while text from the *Glasgow Herald* newspaper did not. Another known source used by Morgan, the

---

[26] For further details, see: https://en.wikipedia.org/wiki/Google_News_Archive (accessed: 21 December 2016).

[27] See: https://news.google.com/newspapers (accessed: 21 December 2016).

[28] https://books.google.co.uk/books/magazines/language/en (accessed 24 February 2017)

Illustrated London News, is fully text searchable for registered users of the Gale News Vault historical newspaper archive.[29]

7. DILIGENT SEARCH: USING CONTENT BASED IMAGE RETRIEVAL (CBIR)[30]

While text-based searching using online search engines will be a familiar concept to researchers, searching for images by using images, is a more recent development. When faced with an image with no caption or clue to its context, Content Based Image Retrieval (CBIR) – also known as reverse image search and search by visual similarity[31] or image retrieval and image mining[32] – is an attractive and easy-to-use research option. It is one which would align well with the diligent search needs of the type of material (de-contextualised images) used in this project. As it is a more recently-developed and lesser used search technique it bears closer examination here.

When a user searches for images using a text-based query, search systems use metadata from the image, text surrounding the image and text in hyperlinks linking to the image,[33] to identify relevant images. CBIR works by indexing images based on their visual content 'such as colors, textures, shapes and regions'[34] to match them with images available online. Typically, a user can upload their own image or use the link for an existing online image as the source image for a search. CBIR technology has traditionally been used by commercial organisations or photographers to identify unauthorised use of their images, but is now being used more widely. The IPO acknowledged the potential usefulness of these tools by including web-based search tools for images in its diligent search guidelines. The IPO guidelines only refer to TinEye (www.tineye.com) and PicScout (www.picscout.com) but other sites are available, such as Image Raider (www.imageraider.com), and an image search function is also embedded in various web browsers. For example, Google has offered a reverse image search function since June 2011, allowing you to upload an image to be compared to visually similar images. Similarly, Bing also offers an Image Match function.

TinEye is free and can be used without registration while PicScout offers a free three-month trial period following registration, features that likely influenced their inclusion

---

[29] See: http://gale.cengage.co.uk/product-highlights/history/illustrated-london-news.aspx (accessed: 21 December 2016).

[30] This section is based on an extract from Patterson, K (2016), *Can I Just Google That? Orphan Works and Image Recognition Tools*, in Andrea Wallace and Ronan Deazley, eds, Display At Your Own Risk: An experimental exhibition of digital cultural heritage, 2016, http://displayatyourownrisk.org/patterson/ (accessed 22 November 2016)

[31] Marques, O (2016), *Visual Information Retrieval: The State of the Art*, IT Professional, vol. 18, no. , pp. 7-9, July-Aug. 2016, doi:10.1109/MITP.2016.70, available at https://www.computer.org/csdl/mags/it/2016/04/mit2016040007-abs.html (accessed 14 December 2016)

[32] Girija O K & M Sudheep Elayidom (2015), *Overview of Image Retrieval Techniques*, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), Vol.4, Special Issue 1, June 2015 , available at http://www.ijarcce.com/upload/2015/si/icrtcc-15/IJARCCE%2019.pdf (accessed 16 November 2016), Section 1

[33] Niuwenhuysen, P (2013), *Search by Image through the WWW: an Additional Tool for Information Retrieval,* published in proceedings of the international conference on Asia-Pacific Library and Information Education and Practices = A-LIEP 2013 http://megaslides.es/doc/4053174/search-by-image-through-the-www--paul-nieuwenhuysen (accessed 21 November 2016)

[34] Niuwenhuysen, P (2013), *Search by Image through the WWW: an Additional Tool for Information Retrieval*

in the IPO's Guidelines.[35] For this reason, they are attractive to the user who is only searching a few images and doesn't wish to sign up to a site or have to pay. TinEye is free for non-commercial users and includes extensions that allow for easy searching in a web browser toolbar. The PicScout Platform is aimed at commercial users. Their search tool is designed to 'enable image buyers to identify and license the images they'd like to use,' and they have '200 million owner-contributed image fingerprints.'[36] As a subsidiary of Getty Images, PicScout would seem an obvious choice when searching for the type of commercial photography that features so heavily in the scrapbooks.

Image Raider relies on Google, Bing and Yandex to get results. It offers a long-term image monitoring service and allows the user to run multiple searches concurrently, features attractive to photographers who wish to monitor potential copyright violations of their work. It uses a credit model, where users can purchase credits or earn credits by tweeting about the site.

Cultural institutions engaging in diligent search activity will rightly be concerned about copyright and image security when uploading images from their collection, and this will no doubt influence their choice of search tool. On this issue, different CBIR systems adopt different approaches. For example, Google's Help Forum states: 'When you search using an image, any images or URLs that you upload will be stored by Google. Google only uses these images and URLs to make our products and services better.'[37] This vague statement will certainly be discouraging to some potential users of the service, who would not want their images retained by Google, particularly in the case of a mass digitisation project. However, as the world's most popular search engine with a global share of 75.2%,[38] it seems a problematic omission from the IPO's guidelines. Would a returning rightholder be satisfied that a diligent search had been carried out without the use of Google?

In contrast, TinEye have a clearer, more acceptable policy: 'Images uploaded to TinEye are not added to the search index, nor are they made accessible to other users. Copyright for all images submitted to TinEye remains with the original owner/author.' [39] Search images submitted to TinEye by unregistered users are automatically discarded after 72 hours, and links to these searches will stop working after 72 hours, unless a registered user happens to save the same image.

Bing's privacy statement does not specifically mention what happens to images,[40] and the Project Officer was unable to find information relating to this issue on PicScout or Image Raider.

---

[35] PicScout was able to be used for free and without registration at the time the UK IPO's Guidelines were last reviewed (September 2016) and also at the time the article on which this section is based, was published (June 2016)

[36] The PicScout FAQs from which this information was taken are now no longer available (checked 16 December 2016) but previously found at http://www.picscout.com/about-us/faqs/ (accessed 8 April 2016)

[37] Google Search Help available at https://support.google.com/websearch/answer/1325808?hl=en (accessed 14 December 2016)

[38] Desktop search engine market share statistics available at https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0 (accessed 14 December 2016)

[39] TinEye FAQs on image uploading available at http://www.tineye.com/faq#uploading (accessed 14 December 2016)

[40] Bing details available as part of Microsoft's privacy statement available at https://privacy.microsoft.com/en-gb/privacystatement/ (accessed 14 December 2016)

So how useful are these search tools? The two IPO recommended tools were tested alongside Google Images, as the leading search engine. The results, when searching for orphan images from the scrapbooks, were variable, especially when dealing with partial or cropped images. Within the scrapbooks, Morgan often cropped down images from their original state in newspapers, magazines and books. These irregular-shaped items tend to decrease the likelihood of an uploaded image search yielding beneficial results, although identification of partial images is still possible.

One example of a successful search is this image of an oil painting (Image 5), taken from Scrapbook 12. Despite the fact that Morgan had cropped the image, Google Images and Tineye were both able to point to sources to identify the cutting showing the

Image 5: Villa Doria Pamphili, Rome (Souvenir d'une Villa)

centre third section of the oil painting *Villa Doria Pamphili, Rome (Souvenir d'une Villa)* 1838-39 by Alexandre Gabriel Decamps (1803-60). Naturally, the key to the success of the search tools is the fact that Decamps' painting can be found on multiple websites. The more ubiquitous the image is online, the greater the chance of identifying it. PicScout, however, was unable to identify the painting.

In general, the kind of hit rate you can expect to get from image search will vary. In a further test of these tools, two pages were selected at random from Scrapbook 12, incorporating a total of 14 viable images. From this small sample, Google provided the best results (identifying two images), followed by TinEye (one), with PicScout unable to provide anything at all. An example of one of the images found by both TinEye and Google is a crop of an animal that appears on a tapestry from the Middle Ages (Images 6 and 7).

Image 6: Cropped image identified as a section from a tapestry by Google and TinEye

Image 7: Original tapestry from which cropped image taken
(www.tchevalier.com/unicorn/tapestries/sight.html)

These findings echo the results of other researchers. Kirton and Terras compared results from TinEye and Google Images in their Reverse Image Lookup study investigating re-use of images from The National Gallery, London.[41] They note that Google produced a 'significantly larger number of results,' giving the example of the painting *Whistlejacket* by George Stubbs which had 109 results on TinEye and 271 on Google.[42] Paul Nieuewnhuysen's study offers similar findings, taking as an example images a set of publicly-available photographs available on a central university server for several years.[43] Nieuewnhuysen submitted nine images to TinEye and Google: Google found six but TinEye found only three.[44] Further searches led to the conclusion that 'search by image for duplicate images functions with an efficiency that is highly variable from case to case.'[45]

Kirton and Terras note that in addition to Google having a much larger database than TinEye (over 10 billion images versus TinEye's just over two billion at the time of their study), they found that Google's database was more up to date.[46] Moreover, as there was little crossover between the results found by Google and TinEye, each system processed the search and presented results in a different way, with TinEye presenting results in a more straightforward and transparent manner and Google constantly readjusting search results as the user moves through pages.[47] These differences help to explain the disparity in search results, and Kirton and Terras argue that the vastly different results 'undermine the use of these tools for anything but a guide as to how to understand image reuse.'[48] Nieuewnhuysen's overview of online image searching reached similar conclusions about the lack of transparency of CBIR systems and the lack of knowledge about their functionalities.[49]

Reverse image search technology can certainly be beneficial to cultural heritage institutions, if used with care. Image recognition tools have a role to play in helping identify any very 'obvious' works which are still in copyright, i.e., those which are usually by well-known creators. Works that are in the public domain are perhaps more likely to be found as they present less risk for users to use and share online; in turn, they are more likely to feature on multiple websites increasing the likelihood of detection. Of course, simply finding an image may not answer the copyright questions you have about the work, but it is a starting point. One example of a useful outcome from the scrapbook sample came from the image search of an advert that originally featured in *The New Statesman*. One result identified the particular issue in which it originally featured as containing spoof publisher adverts and in-jokes, which was not evident when the advert was removed from that context. This demonstrates an additional use that image

---

[41] Kirton, I and Terras, M (2013), *Where Do Images of Art Go Once They Go Online? A Reverse Image Lookup Study to Assess the Dissemination of Digitized Cultural Heritage.* In Museums and the Web 2013, N. Proctor & R. Cherry (eds). Silver Spring, MD: Museums and the Web. Published March 7, 2013. Available at http://mw2013.museumsandtheweb.com/paper/where-do-images-of-art-go-once-they-go-online-a-reverse-image-lookup-study-to-assess-the-dissemination-of-digitized-cultural-heritage/ (accessed 14 December 2016)

[42] Ibid.

[43] Niuwenhuysen (2013).

[44] Ibid.

[45] Ibid.

[46] Kirton and Terras (2013), section: 'TinEye versus Google Images Search'.

[47] Ibid.

[48] Ibid.

[49] Niuwenhuysen (2013).

recognition tools might offer beyond the identification of a possible rightholder: they can help cultural heritage organisations contextualise and research the material within their collections.

There are also practical considerations to bear in mind when using these tools. Preparing images to upload for search may involve considerable effort that is not scalable when engaging in a mass digitisation project. For example, we estimate that all 16 volumes of the Edwin Morgan scrapbooks contain 41,472 orphan works requiring diligent search and, from the sample, 56% of items are artworks and photographs from published sources, leading to a large amount of potential administration time for carrying out checks using CBIR tools. Moreover, institutions should also consider which tool is the most appropriate tool for its needs. Although Google was the most likely to provide results, individuals and organisations will have understandable reservations about uploading large amounts of images to Google due to their policy of storing images for their own use. Equally, other sites with unclear or unstated policies about image storage may also be an unattractive choice for users.
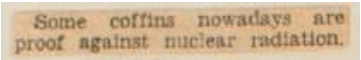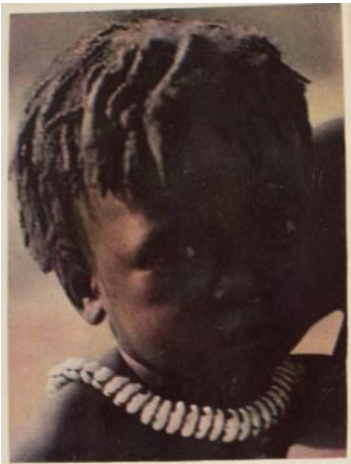
Ultimately, in the case of the scrapbooks, the nature of the de-contextualised works means that in some cases, CBIR systems form a significant way of conducting an-IPO approved diligent search. The technology is still relatively new and results are variable; it should be used with this in mind. However, as the technology is continually developing and improving, it seems likely that the usefulness of image recognition tools for cultural heritage institutions engaging in digitisation and rights clearance activities will only increase in the future.
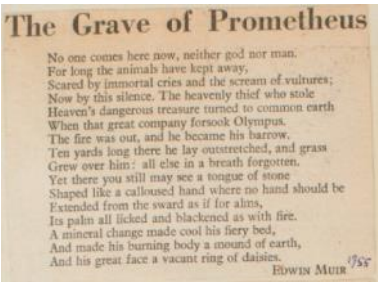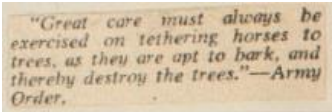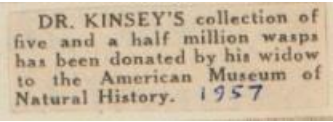
8. DILIGENT SEARCH EXAMPLE: PAGE FROM SCRAPBOOK 12

In this section we present details of the search activity carried out in relation to one complete page from our sample. Table 5 provides narrative detail of the search history for each cutting, as well as the time spent on each search, as well as the total period of time over which each search was conducted. We consider each of these examples to constitute a good faith diligent search.

Image 8: Page 2245 from Scrapbook 12

IMAGE	SEARCH HISTORY	TIME SPENT	TIME PERIOD
Some coffins nowadays are proof against nuclear radiation.	Tried web searches for a match to sentence in cutting. No matches.	10 mins	1 day
	Cropped image from page so it could be checked by Image Recognition Tools. PicScout and TinEye did not recognise, Google Images recognised it as part of a cover from a 1954 French magazine called Réalités from May 1954 with title strapline 'le village le plus secret de l'afrique'. Magazine not available online or in the Glasgow University library. World Cat showed that it is available in the [National Library of Scotland in Edinburgh](). Further online research showed that an English edition of the magazine was also produced. Searched for images from a couple of photographers who were named as frequent contributors - Jean-Philippe Charbonnier and Édouard Boubat. In an image search filtered by colour images, none similar came up. No results for image search of full cover of magazine. Time spent: 40 mins.

Further web research and looking at names of photo agencies to approach – decided to try an orphan works search through BAPLA (25 mins). Emails with Getty and Camera Press. Reply from LAT Photo (negative) Camera Press (negative) PA Images Getty (negative) Lee Miller Archive (negative) (35 mins) Time spent: 1 hour.

A copy of the magazine was located in the National Library of Scotland in Edinburgh. After applying online for a library card, a letter was delivered to my home address with the card number so the magazine could be ordered to the reading room for my arrival at the library (36 mins). Following a return train trip to Edinburgh (3 hours 40 mins for travel time and time spent at library), I did not find the photograph I was looking for in the magazine. The May 1954 issue of the magazine in the Library had a different cover. It was the English version so it is possible that Morgan used the French version and it may have had a different cover. Or, it came from a different issue of the magazine altogether (I checked January to June 1954). On the positive side, I did discover several other images from the magazines that were used in the Scrapbook. Train fare: £24.90 | 5 hours 56 mins | Dec.2015 – Sep.2016 |

CREATe

	After discovering dates for Edwin Muir were 1887-1959, I found him online on the [Faber website](). Initial web research and then online application for permissions, jointly with another work (40 mins) No response after stated 6-8 week reply period. Follow up correspondence (10 mins) but further follow up was required several month later as some questions remained unanswered and issue took around seven months to be resolved. Finally, Faber concluded that, 'I'm afraid we don't have contact information for the Estate of Edwin Muir. We can, however, suggest that you look into Fair Dealing in case you feel that the usage falls under that particular remit.' (10 mins).	60 mins	Nov. 2015 – Aug. 2016
	Tried web searches for a match to the sentence. No matches.	10 mins	1 day
	Tried a web searches for a match to the sentence. No matches.	10 mins	1 day
	Decided item could be deemed not substantial enough to be covered by copyright law. It is cut from a larger image and it is hard to determine the subject of the photograph.	5 mins	1 day
	Cropped image from page but there were no image search matches on Google, TinEye or Picscout.	15 mins	1 day

	Cropped image from page but there were no image search matches on Google, TinEye or Picscout.	15 mins	1 day
	Contact with Newsweek over 3 month period via emails, web contact form and finally via Twitter, with difficulty determining who was able to answer the question on rights. Finally resolved, Newsweek granted rights for us to use the text, they are unable to grant image rights as images are credited to the US Airforce (60 mins) Works created by the US Government are in public domain. Not completely clear if this includes the forces and the US Government may assert copyright for US Gov works used outside of the US: https://www.usa.gov/government-works. Spent time researching US Airforce copyright and general US copyright, emailed military contact details found online over a period of three months but no response. Decided to assume that images were in the public domain and use (70 mins).	2 hrs 10 mins	Nov.2015 – Jan. 2016
	As with the cutting above, decided that credit to US Airforce meant that item was in the public domain and could be used. The caption was deemed to be sufficiently factual so as to not to attract copyright protected.	10 mins	1 day
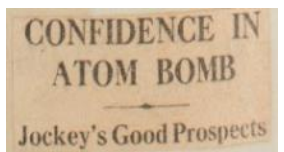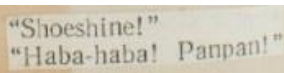
	Cropped image from page but there were no image search matches on Google, TinEye or Picscout.	15 mins	1 day
	Cropped image from page but there were no image search matches on Google, TinEye or Picscout.	15 mins	1 day
	Tried web searches for a match to words in cutting. No matches	10 mins	1 day
	Tried web searches for a match to words in cutting. No matches	10 mins	1 day

Table 5: Details of diligent search activity for p.2245

9. SELECTING WORKS FOR AN OWLS APPLICATION

To explore the parameters of OWLS, we made an application using different types of work: two cartoons, a poem, a text cutting from a magazine and an original black and white photograph. Predominantly, these were works where a significant amount of time had been spent in trying to find the rightholder, as the name of a potential rightholder was included in four of the five items (see Table 6). The names of the cartoonists and poet were provided on the cuttings and although the name of the author of the magazine text was unknown, the source publication was given. A diligent search was conducted for each of these items relying on the IPO guidelines.

The remaining item was an original photographic work (Image 9), rather than one cut from a magazine or other source. We believe it to be a purchased black and white studio photograph, contemporary with the creation of the scrapbooks; however, Morgan provides absolutely

Image 9: Original black and white photograph submitted for OWLS application

no information about its origin. With this studio photograph, we were interested to explore the IPO's response to the use of image recognition tools as the primary means of diligent search in an application made to the OWLS scheme. Although they are an IPO-approved method for diligent search, the results can be variable, as previously discussed. Google, PicScout and TinEye were used to search for the photograph, with no results; an application was then submitted to the OWLS scheme on the basis of just those three searches. The IPO's response was that the requirements of the scheme would be satisfied by a further search of three additional sources: the Association of Photographers (AOP), British Association of Picture Libraries (BAPLA) and British Institute of Professional Photographers (BIPP). This involved sending an email to each contact and did not result in identification of the work: the AOP undertook to forward details on to their members and Board of Directors and notify us of identification was made but no further contact was received, BAPLA sent out details to their image suppliers but there was no positive response and BIPP could not identify but would keep the details on file. The total time spent on the search was 25 minutes, over a period of one day. The work was deemed to be an orphan. Again, the IPO agreed.

This result should be encouraging to cultural heritage institutions who intend to apply to OWLS, to know that a diligent search carried out using CBIR tools can form a significant part of their application.[50]

10. MAKING AN OWLS APPLICATION [51]

This section explores the mechanics of the diligent search and application process for OWLS, focussing on the cartoon by Paton, looking at the diligent search undertaken and the process from making the application to the licence being granted. Following the guidance provided by the IPO, the Project Officer undertook a diligent search for the

---

[50] This section is based on an extract from Patterson, K (2016), *Can I Just Google That? Orphan Works and Image Recognition Tools*, section IRTS and Diligent Search

[51] Section based on Stobo, V., Patterson, K., and Erickson, K. (2017) Forthcoming.

rightholder of this cartoon by using official sources listed in the guidelines,[52] in addition to other sources as appropriate. While comprehensive, the guidelines do not claim to be complete, but provide 'details on the relevant sources that applicants must consult and provides a non exhaustive list of additional sources.'[53] Thirty minutes of the diligent search consisted of general web research, including a search of the existing UK orphan works registry and contacting the British Cartoon Archive (BCA) for information on Paton. The BCA replied after a number of days with a web-link to a different cartoon by Paton that had appeared in the magazine Parade, but could provide no other information. The BCA are not listed on the guidelines but were identified during web searching as a possible source of relevant information.

The IPO guidelines for Still Visual Art list eight 'Illustration Associations' which were of particular note for this type of work. Of these, four were disregarded immediately as not relevant, as they related specifically to railways, sculpture, architecture and medical illustrations. The other four were the Professional Cartoonists Organisation (PCO), the Comic Creators Guild (CCG), the Cartoonists Club of Great Britain (CCGB) and the Association of Illustrators (AOI). The CCG is primarily concerned with strip cartoons so was not appropriate. Paton was not a listed member on the sites of either the PCO or the CCGB, and the PCO were unable able to help with research enquiries. An additional step taken beyond the boundaries of the guidelines was to post on the forum of the CCGB website, which is reasonably active. Twenty-five minutes spent registering and engaging with the CCGB forum resulted in a reply but it did not identify the artist.[54] The Project Officer spent fifteen minutes emailing both the AOI and Punch (not included in the guideline but a known source of cartoons of this type) but received a non-response and negative response, respectively. In total, one hour and 10 minutes was spent on diligent search for the item, with no positive results.

Image 10: Cartoon by Paton submitted for OWLS application

Additional time was spent on the administrative task of applying for a licence to make use of the orphan work. The OWLS application process takes place entirely online. It requires comprehensive information about the work but for the purposes of this section, and in the interest of brevity, we provide a condensed account of the process only.[55]

In completing an application, the first problematic issue encountered concerned providing a title for the work: where a work has no obvious title, how can applicants be assured that they are choosing a title which would enable the work to easily be found by other users? For example, a portrait photograph of a soldier could be titled as 'Portrait of a soldier', or by someone with knowledge of regimental badges as 'Portrait of a soldier

---

[52] UK IPO Guidelines available at https://www.gov.uk/government/publications/orphan-works-diligent-search-guidance-for-applicants (accessed 22 November 2016)

[53] As found in the introductions to the Guidance relating to still visual art, film and sound, and literary works:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/474401/orphan-works-literary-works.pdf (accessed 22 November 2016)

[54] The reply is available at: http://www.ccgb.org.uk/q_and_a_forum/simpleforum_pro.cgi?fid=01 (accessed 14 December 2016)

[55] See appendix in Stobo, V., Patterson, K., and Erickson, K. (2017) Forthcoming

from X regiment,' a much more specific title. For most published sources this is unlikely to cause significant problems, although there could potentially be issues with items that are published under different titles in different jurisdictions. It is possible to note that the item has no title, and to provide a description instead. Licensees should be aware that rightholders and other users might search the register periodically, so any title given for a work should include information likely to be used as part of a keyword search. If titles given to works in archive or museum catalogues are unlikely to satisfy this requirement, some consideration should be given to the time and effort that devising appropriate titles for the application process might generate at this stage of the overall rights management process.[56]

During the application, it was also necessary to make a number of assumptions about the work in order to proceed. The most difficult assumption to make was whether to identify Paton, the cartoonist that created the work, as the rightholder in the work. We know nothing about the publication the work was taken from, and nothing about Paton. It could be the case that the publisher holds the rights to the work. Indeed, this was a recurring issue for the scrapbooks as whole given the huge amount of the material contained in the volumes taken from newspapers published in the 1950s. The work of freelance journalists is specifically mentioned in both the 1911 and the 1956 Copyright Act, with copyright being retained by the writer when working in a freelance capacity. [57] However, without access to employment records, it is difficult to gauge whether the journalist (or in this case, cartoonist) was working under a contract of employment or whether they were a freelance worker which would impact who owned the rights in question.

One also has to indicate whether the application relates to a commercial or non-commercial purpose. On this point, the definition of commercial use employed by the IPO is worth considering. According to the IPO, their definition 'has been chosen to reflect the practice of licensors. For the avoidance of doubt, it is not intended as a definition in UK or European copyright legislation'. They continue:

> Commercial use covers any uses (including by individuals as well as organisations) that make money from the work – such as selling copies of the work or charging directly for access to it. As well as activities that generate revenue, such as merchandising or selling copies of a publication, commercial use would also cover any other uses that are commercial in nature, such as any use in commercial advertising, marketing or promotion activities. This applies equally to not-for-profit organisations.[58]

---

[56] Archivists, librarians and curators are generally very adept at creating titles for works in their collections, through cataloguing, but it may often be that a title in a catalogue isn't specific enough - they may be relying on an identifying number rather than a descriptive title. For example, to identify individual cuttings within the scrapbooks, the project officer had to devise a numbering scheme. This meant extra time had to be allotted to the creation of descriptive titles when the applications were submitted.

[57] See the Copyright 1956 Act Part, 1 s.4(2) available at http://www.legislation.gov.uk/ukpga/1956/74/pdfs/ukpga_19560074_en.pdf (accessed 14 December 2016), and also the 1911 Copyright Act available at http://www.legislation.gov.uk/ukpga/1911/46/pdfs/ukpga_19110046_en.pdf (accessed 14 December 2016).

[58] Intellectual Property Office (2015) *Orphan Works Licensing Scheme overview for applicants*, 2.34 - 2.38, available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/518251/Orphan_Works_Licensing_Scheme_Overview_for_Applicants.pdf (accessed 17 November 2016).

The IPO's definition of commercial use could affect even very small scale endeavours. Take the example of a local history society wishing to use four or five orphan images in a booklet with a small print-run. Even if they don't intend to profit from the publication, but only want to cover printing costs by charging a small fee, the society would still be charged the same commercial rate as a much larger publisher to make use of the orphan works. Indeed, the cost of the licence could make the planned publication unfeasible, unless they are willing to raise the price per copy.[59]

A final consideration for institutional users of the scheme is the fact that debit and credit cards are the only accepted method of payment, which is due in two stages: an application fee to start the process and a licence fee once the IPO accepts the application and grants a licence. For smaller applications containing only a few works, it may be difficult for local authorities and large institutions to pay a fee of as little as ten pence (£0.10) for the non-commercial use of a single work, without the option of raising an invoice or paying an additional card-handling charge.

The application process, in its entirety, took one hour and 10 minutes. Diligent search took the same amount of time, providing a total of two hours and 20 minutes spent on the single Paton cartoon. Table 6 presents these results alongside the other works for which an application was made to the OWLS scheme. A licence was granted for the non-commercial use of all five works.

---

[59] This mirrors the EU Directive on Re-use of Public Sector Information, which requires that all users be charged the same amount for re-using public sector information.

NAME OF WORK	TIME SPENT ON DILIGENT SEARCH (DS)	TIME COMPLETING DS CHECKLIST	TIME ON APPLICATION	TIME ON DS REQUESTED BY IPO	TOTAL TIME	APPLICATION FEE (AVERAGE)	COMMERCIAL COST	TERMS OF USE	NON-COMM-ERCIAL COST
Paton Cartoon	70 mins	20 mins	50 mins	10 mins	140 min	£2.67	£75.12 +VAT	Reproduction in a published book (not a textbook), Inside the book, 1/16 page or less, one year or less, 1000 copies or less	£0.10
Appleton Poem	85 mins	15 mins	10 mins	5 mins	110 min	£2.67	£96.43 + VAT	Reproduced as a book, up to 1,000 copies, sale price of £10	£0.10
Giovannetti Cartoon	100 mins	25 mins	20 mins	15 mins	160 min	£2.67	£89.15 + VAT	Reproduction in a published book (not a textbook), Inside the book, 1/4 page or less, one year or less, 1000 copies or less	£0.10
Doubt Magazine cutting	180 mins (60 online/120 library)	15 mins	20 mins	15 mins	230 min	£2.67	£56.25 + VAT	Reproduction in a published book (not a textbook), Inside the book, 100 words or less, one year or less, 1000 copies or less	£0.10
B&W photo of man	10 mins	15 mins	10 mins	15 mins	50 min	£2.67	£77.13 + VAT	Reproduction in a published book (not a textbook), Inside the book, 1/8 page or less, one year or less, 1000 copies or less	£0.10
Total:									**690 min**
Average:									**138 min**

Table 6: Results of diligent search for works on UK IPO application

11. SELECTING WORKS FOR AN EUIPO REGISTER APPLICATION

Selecting works for the EUIPO register, there is a more limited choice as artistic works are not eligible. That is, works such as the original photograph registered under OWLS are ineligible. However, embedded artistic works do fall within the scope of the Directive,[60] and these form the majority of the artistic works contained in the scrapbooks. And yet, within the context of the scrapbooks these embedded artistic works pose their own problems. In their original context – the newspaper or magazine from which they were taken – these cuttings are indeed embedded. But do they cease to qualify as an embedded work when neatly cut from those newspapers and magazines to be pasted into a scrapbook alongside other freestanding artistic works such as the black and white photograph previously discussed? Would a returning rightholder regard the cutting of what was an embedded work to be no longer embedded? Possibly not. Indeed, often with the images included in the scrapbooks it can be difficult to distinguish between an original photograph and a reproduction taken from a magazine or a book. In any event, for the purpose of this project we interpreted the Directive to apply to embedded artistic works even when those works had been removed from their original publication.

To evaluate the time and resource costs involved in using the EUIPO orphan works database and exception, five works were chosen: three news cuttings and two embedded photographic images, one from a magazine and one from a newspaper. These items were chosen for their similarity in length and provenance to the works used in the OWLS exercise, but they were unique in order to simulate a first-time rights holder diligent search.

12. MAKING AN APPLICATION TO THE EUIPO REGISTER[61]

We registered five works with the EUIPO register. As with OWLS, registering works on the EUIPO orphan works database takes place online. Before registering works, institutions must register as a 'beneficiary organisation' with the EUIPO. The application process is straightforward, but it took five working days for registration to be confirmed by the EUIPO permitting log-in and registration of works on the database.

Image 11: News cutting submitted to EUIPO for Orphan Works Register

Many of the features noted during the OWLS application apply equally here: instead of creating a title for the work, users can record that the work has no title, and provide a full description instead. Assumptions about rightholders still have to be made: in this case, the project team assumed that the newspaper publisher would hold copyright in the cutting, although

---

[60] Article 1(4) of the EU Directive states: 'This Directive shall also apply to works and other protected subject-matter that are embedded or incorporated in, or constitute an integral part of, the works or phonograms referred to in paragraphs 2 and 3'; that is, although not included in their own right, artistic works printed within the 'books, journals, newspapers, magazines or other writings' listed in Article 2(a), are considered to be embedded works.

[61] This section is based on Stobo, V., Patterson, K., and Erickson, K. (2017) Forthcoming.

there is certainly an argument that the journalist could (or perhaps should) be listed as an additional rightholder (depending on the specifics of the contractual agreement between journalist and employer).

That said, the process of applying to the EUIPO register is less onerous than the UK licensing scheme: the online form is much shorter and the database offers a bulk upload function that simplifies the registration process. EUIPO request that users submit spreadsheets to them before upload: this is simply to check that data can be processed; EUIPO does not audit individual diligent searches. This bulk upload function makes it significantly quicker to register works; although the same information fields required by the web form has to be completed in the spreadsheet, completing the information in a spreadsheet format will be quicker than individual online registration.

However many works one is registering, it is important to record the narrative of all diligent searches, the sources used, the results, as well as documenting how the work is used; moreover, keep those records for at least as long as the work is registered and in use.

Finally, when relying on the Directive, it is important to keep in mind some significant differences between the Directive and OWLS. We have already addressed the issue of free-standing artistic works. The other main differences are discussed in *Orphan Works: The Legal Landscape*; they include: (i) the limited application of the Directive to unpublished works (OWLS applies to all types of work whether published or not); (ii) the need to provide 'fair compensation' to a reappearing orphan work owner under the Directive (under OWLS, a reappearing owner is only entitled to claim the licence fee already been paid under the scheme); and (iii) that the Directive enables non-commercial use only (OWLS permits both commercial and non-commercial use).

Image 12: A magazine photograph submitted to EUIPO

Table 7 presents details of the time and other associated costs when registering our sample of orphan works with EUIPO.

Image 13: A magazine photograph submitted to EUIPO

NAME OF WORK	TIME SPENT ON DILIGENT SEARCH (DS)	TIME SPENT ON APPLICATION	TOTAL TIME SPENT
Newscutting 'Atomic Particle's Travels'	25 mins	17 mins	43 mins
Newscutting – Football Pools winner	20 mins	14 mins	34 mins
B/W newspaper photograph – mushroom cloud	10 mins	16 mins	26 mins
B/W magazine photograph – doorways	10 mins	11 mins	21 mins
Newscutting – BIS takes you there	15 mins	10 mins	25 mins
Total			158 mins
Average			31.6 mins

Table 7: Time and associated costs under the EUIPO scheme

13. CLEARING THE SCRAPBOOKS [62]

Having carried out the data extraction exercise and made applications under OWLS and the EUIPO register, we used this information to estimate the amount of time it would take to clear the 16 volumes of scrapbooks under both orphan works regimes.[63]

We estimated the total number of works in the collection to be 51,480, based on an average of 15 cuttings per page across 3,600 pages in 16 volumes. Table 8 provides an overview of the different types of works included in the sample. Having identified that 69% of the sample were artistic works, the Project Officer conducted further analysis to identify the proportion of works which were (originally) embedded within another work. Only 13 works were found to be standalone artistic works; the remaining 285 works were deemed to be embedded works for the purpose of the Directive.

---

[62] The results presented in this section are taken from Stobo, V., Patterson, K., and Erickson, K. (2017) Forthcoming.

[63] Both regimes would need to be used as the Directive does not extend to free-standing artistic works.

TYPE OF WORK	NO. OF WORKS	PERCENTAGE OF SAMPLE
Artistic Works		
Book Artwork	7	
Book Photograph	10	
Newspaper Artwork	8	
Newspaper Photograph	23	
Periodical Artwork	15	
Periodical Photograph	177	
Original Artwork	11	
Original Photograph	37	
Photograph of a TV broadcast	1	
Ephemera	9	
TOTAL	**298**	**69%**
Text-based works		
Book Text	9	
Newspaper Text	74	
Periodical Text	44	
Original Handwriting	7	
TOTAL	**134**	**31%**
Overall Total	**432**	**100%**

Table 8: Types of works included in the sample

In Table 9 we summarise the rights status of all 432 works in the sample. 52% of the sample were revealed to be orphan works; scaling up for all 16 scrapbooks, this gives an estimate of 26,770 orphan works in total.[64]

---

[64] This number was extracted by calculating the percentages from the number of orphan works within the sample (226, or 52%) by the number of estimated works across the 16 volumes of scrapbooks (51,480). The three works where a response has not been received and the project staff are unsure they have the correct contact details for the identified rightholders, could be considered orphan works according to the definition provided in the CDPA 1988. However, we do not count them as such for the purposes of this analysis.

TYPE OF WORK	NO. OF WORKS	PERCENTAGE OF SAMPLE
Works created by Edwin Morgan	37	9%
Works without or unlikely to attract copyright protection		
Works in Public Domain	22	5.1%
Insubstantial parts of works	84	19%
Ephemera (e.g. adverts and tickets)	9	2.1%
Permission Granted		
Works where permission granted by rightholders (15) for use free of charge (4 works subject to specific instructions)	29	7%
Works where permission granted by rightholders (7) for use dependent on fee (which project staff decided not to pay)	19	4.4%
No response		
No response from rightholders (2), with accurate contact details	2	0.5%
No response from rightholders (3), with uncertain contact details (these could be defined as orphan works)	3	0.7%
Works where publisher was unable to license for online use (advised staff to use an exception instead)	1	0.2%
Definite orphan works (excluding non-responders where address details are uncertain)	226	52%
Total works in sample	**432**	**100%**
Total time spent on diligent search	**1080 hours**	
Equivalent salary cost	**£11,653.20**	

Table 9: Rights status across the sample of 432 works from Scrapbook 12

The total number of estimated orphan works, and the indicative status of works within the scrapbooks, allows us to estimate the likely costs of undertaking diligent search activity and clearing rights for all 16 scrapbooks. If we assume the total number of orphan works to be 26,770, then 94% (or 25,164) can be uploaded to the EUIPO orphan works database and 6% (or 1,606) can be licensed through OWLS.

In relation to the 1,606 works that would have to be licensed through OWLS, assuming a standard salary cost of £10.79 per hour,[65] the total estimated costs are as follows: the

---

[65] We report using the average hourly rate of £10.79, calculated from the most conservative archivist annual salary estimate or £22,443 as reported by the Archives and Records Association (ARA). 'The ARA recommends that the minimum starting salary for recently qualified archivists, archive conservators and records managers is between £22,443 and £38,000'; see: http://www.archives.org.uk/careers/careers-in-archives.html (accessed 22 December 2016).

application and licence fees would amount to £4,448.62; the salary costs and time spent on the rights clearance process would be £39,860.92 and 1.8 years respectively (based on 52 weeks at 40 hours per week).[66] For those works falling within the scope of the Directive (25,164 in total) the average time per work complying with the diligent search requirements and interacting with the online system was 31.6 minutes. This means the salary costs and time spent on the process would be £142,931.52 and 6.4 years, respectively. It is likely, however, that this total could be further reduced by making use of the bulk upload mechanism and grouping works together.

The total cost of using both OWLS and the EUIPO Orphan Works database in tandem to make all orphan works contained in the scrapbooks available online would be £187,241.06 (including application and licence fees, and salary costs) and would take 8.2 years.

14. CONCLUSION

Rather than present conclusions within each section, we collect them together within one project conclusion (available at www.digitisingmorgan.org/conclusion). In the next section of the resource (http://www.digitisingmorgan.org/Risk) we consider risk assessment as an alternative to strict copyright compliance, and outline how we engaged in the process of risk assessing the digitisation of our 10% sample of Scrapbook 12 from Edwin Morgan's Scrapbooks.

---

[66] These figures are based on the application and licence fees required by the UK IPO, which we have calculated at £2.67 per work and £0.10 per work respectively. This is based on the assumption that applications will be made in batches of 30 works at a cost of £80 in application fees (this is the maximum amount allowed in a single application). The non-commercial licence fee of £0.10 is the standard amount charged under the scheme. The average time to complete a diligent search in line with the expectations of the IPO, and engage with the application process, was 138 minutes; at a standard salary rate of £10.79 per hour, this generates salary costs of £24.82 per work.