

Copyright and Data Protection in relation to literary archives and manuscripts: GLAM meeting, 10 September 2007

I. Introduction and Background

This document sets out some reasons why GLAM members believe that copyright and data protection raise particular, and often problematic, issues in relation to literary archives and manuscripts. These are followed by specific questions on copyright and data protection issues submitted by GLAM members, as well as a number of case studies – some of which are merely observations, and some of which include further questions for our speakers.

By way of background, the following, taken from GLAM's website, gives some indication of the kind of material which forms the focus of GLAM's work:

GLAM defines 'literary' as creative writing in all genres, whether or not intended for publication, including:

- Poetry
- The novel
- Other forms of fiction (including short stories and novellas)
- Other prose writing, such as essays and letters by or relating to 'literary' figures
- Writing for drama, in the theatre, or for radio, television and film
- Life writing, including literary biography, autobiography and self-representation
- The writing of criticism (both theoretical and practical) relating to creative writing, including editing, reviewing and the histories of literature
- The process of publishing or otherwise supporting the production and dissemination of literature

GLAM recognises that literature has no boundaries of language, geography or ethnicity. For practical purposes GLAM is likely to focus its interest on Literatures in English, and the other languages which are indigenous to, or have taken root within, the archipelago of the British Isles (including England, Wales, Scotland, and the whole of Ireland). However, it also takes account of: literatures in other languages which have been translated into English; and foreign language material represented in literary collections held by British and Irish institutions.

The chronological remit of GLAM spans from the end of the Middle Ages to the present. However, the work of the group is likely to focus on the modern period, particularly the 19th-21st centuries.

GLAM recognises the wealth and diversity of literary collections, and therefore does not limit its definition of literary archives to specific formats or record types. Typically, any literary archive may be comprised of some or all of the following:

Literary manuscripts: These are documents (in a range of formats) which relate to all phases of the genetic history of a literary work through the stages of its life cycle - from the avant-texte stage, through the textualization stage, to the post-text stage. They include documents created during the following phases of work:

- Pre-compositional phase: notes, sketches, drawings, work plans, notebooks, marginalia, annotations.
- Compositional phase: rough drafts, reworkings, corrected fair copies.
- Pre-publication phase: reworkings of manuscript and definitive manuscript.
- Post-compositional phase: author's proof copies, editor's proofs, collated proofs, other production records.
- Post-publication phase: author's annotated editions of a published work.

Letters: These can be personal (family, friends, lovers), literary and cultural (other writers, artists, musicians, film makers, performers, critics, academics, reviewers), or business-related (agents, publishers, literary organizations and institutions, gas board, bank).

Diaries: These can be personal or engagement diaries.

Audio-Visual Material: This can include photographs, films, videos, tape recordings, CDs, drawings.

E-Records: These can be software files, emails, websites, floppy disks.

Printed Sources: These can be newscuttings, annotated books, journals, magazines.

Legal Documents and economic records: These can be contracts, court proceedings, cheque books, account books, bank statements.

Objects: These might include pens, locks of hair, typewriters, writing desks.

2. Do literary archives and manuscripts constitute a 'special case' in relation to copyright and DPA?

GLAM members believe that there are some issues raised by literary archives and manuscripts which make administering copyright and data protection different, and in some ways more problematic, than it is in relation to other types of archive.

2.1 Copyright

- While this applies to all archives, it has greater implications for literary papers – particularly those which contain previously unpublished material. In the context of successive drafts of a work with subtle differences underlying a published

version, it is difficult to establish when a manuscript should or should not be regarded as unpublished.

- The prohibition on photocopying for researchers any part of a copyright literary manuscript which had been published before deposit in a repository can be problematic – particularly for those institutions which collect the archives of publishing houses. It also seems anomalous in light of the fact that entire unpublished manuscripts can be copied for private study and non-commercial research if they were *not* published before deposit; if anything, unpublished material is more vulnerable to plagiarism and misuse than the manuscripts of already-published works.
- Literary archives are often seen as good candidates for digitisation: the papers of a single writer are usually widely dispersed, and many literary collections are purchased by overseas institutions; digitisation has been suggested as a means of overcoming this dispersal, providing full access to a writer's output, and rendering physical location irrelevant. However, digitising the archive of a single writer is likely to involve clearing thousands of copyrights, which would be prohibitive for most repositories.
- Many of the manuscripts and papers of British and Irish writers have been purchased by US institutions, and problems can arise as a result of differences in national copyright legislation, which can lead to the papers of a single writer being out of copyright in one country (and therefore free for researchers to publish) but still restricted in another.
- The monetary value of literary papers means that in some cases prize items are 'cherry picked' from an archive, sold into private hands, and replaced by photocopies or other surrogates of the originals (often without copyright permission having been sought first); this raises copyright problems for the repository which subsequently receives the bulk of the archive, including the infringing copies.
- Literary copyrights may have, or be thought to have, unusually high value. For this and other reasons (e.g. concern over a writer's reputation), dealing with literary estates can be difficult; some estates have refused copyright permission for the use of unpublished papers, or refused access to material except to an 'authorised' biographer.
- The importance of copyright in relation to literary papers led to the establishment of the WATCH database by the Harry Ransom Humanities Research Center (University of Texas) and the University of Reading Library. Literary archivists are fortunate to benefit from this resource which is invaluable for tracing literary copyright holders, and GLAM fully supports the WATCH initiative.

2.2 DPA

- Literary archives typically contain much personal data, and frequently also sensitive personal data; this applies both to personal archives (e.g. the papers of an individual writer) and to institutional papers (e.g. small literary magazines and publishing houses, where publishers and writers are often also friends and move in the same social circles). For example, a single archive held by a GLAM member repository includes papers relating to: sexuality; adultery; illegitimacy; mental and physical illness; libel; suicide attempts; criminal convictions; alcoholism; drug abuse; and more.
- Administering the DPA in relation to literary papers can be very labour intensive. While standard guidelines and codes of practice are relatively straightforward to apply to institutional records, or uniform records series, they are often more difficult to apply to literary papers which can be unstructured and highly varied in content. Sensitive material can be buried deep within an archive; inevitably this means that every single document needs to be read before it can be safely issued to a researcher in the reading room; the alternative is imposing blanket closures on any material relating to living individuals, which conflicts, in many cases, with FOI. There is also little uniformity in the way that DPA is implemented between different institutions.
- Writers often choose to make their private lives public through their creative writing; this blurs the distinction between the public and the private, and makes closure decisions more problematic.
- Often literary researchers are those who are most intent on uncovering the most personal aspects of the lives of their subjects; this does not just apply to literary biographers – in literary research all written documents potentially contain information relating to the evolution of finished literary texts.
- Determining what might cause ‘substantial damage and distress’ to a particular individual is a highly subjective and unpredictable process.

3. Copyright and Literary Papers

3.1 Questions

3.1.1 *Comparative duration of copyright in the UK and USA*

Since 2003, duration of copyright in manuscripts in the USA has been reduced to 70 years from the death of the author, whilst Britain still has the transitional arrangements which will continue until 1 January 2040. This means that a D. H. Lawrence manuscript in the University of Texas is considered by them to be in the public domain, whilst a part of the same manuscript sequence in the University of Nottingham is in copyright

until 2040. Is Tim Padfield aware of any developing working practices to deal with this anomaly? Does he have any reflections on this disparity?

3.1.2 Exhibition of literary manuscripts

Literary papers have great popular appeal for use in exhibitions. Tim Padfield's book suggests that 'the unauthorized public exhibition of copyright literary, dramatic and musical works is an infringement of copyright, but does not count as publication...The work may not, of course, be reproduced in a catalogue of an exhibition without permission, since that involves issue to the public'. Does this imply that we should be seeking copyright permission for every original copyright item we wish to display in an exhibition (even if we do not wish to reproduce it in a catalogue)? While we would always seek permission from living individuals before displaying one of their manuscripts or letters (for Data Protection reasons), we would not usually think of seeking permission from a literary estate for displaying an item that was (for example) 200 years old.

3.1.3 Digital preservation and the Gowers Review

Increasingly, literary manuscripts are being created in digital form, and writers are carrying out their correspondence via email. Repositories are therefore having to address the issue of digital preservation, which requires the creation of multiple copies of each record, and sometimes also involves modifying the original digital object in some way, e.g. saving multiple copies of a bitstream, refreshing the bitstream to new media, and migrating to different formats. Currently, all of these activities infringe copyright. The Gowers Review recommends that UK law should be amended to allow libraries to copy the master copy of any digital work for 'archival purposes' and to make further copies or format shift to avoid obsolescence. Does Tim Padfield have any information about whether the recommendations of this review are likely to be adopted by Government and become part of UK copyright legislation?

3.1.4 Access in the digital environment

The latest edition of Tim Padfield's book provides valuable guidance on copyright in the electronic environment, setting out what is and is not permitted in terms of copying for private study and non-commercial research. However, much of this guidance focuses on material which is classed as 'published', and is subject to the same rules as hard-copy published material, e.g. printing out/copying to disk of 5% of a work is permitted, whereas copying the whole of an electronic work is not. In the paper environment, we are permitted to copy the whole of an unpublished work for a researcher. Much of the digital material which forms part of literary archives will be unpublished, e.g. notes for, and early drafts of, literary works, and personal documents. Does the guidance on digital material given in Tim Padfield's volume also apply to unpublished material or should we apply the standard hard-copy guidelines for unpublished papers?

3.1.5 Email

Curators of literary archives are likely to be dealing with extensive quantities of email in future as publishing houses and individual authors increasingly conduct their correspondence electronically. Obviously the strictures of the DPA mean that most email acquired now will have to be closed to researchers for lengthy periods, simply because of the impossibility of reading the entire contents of an email directory before making it available. However, it would be helpful to have the copyright situation in email clarified fully. Tim Padfield's book suggests that the sender owns copyright in an email sent to a single recipient, whereas he/she can only claim moral rights over a message sent to multiple recipients. Should we interpret this literally, e.g. the sender does not own copyright in a message sent to two or more individuals? In which case who does?

3.1.6 Inter-generational transfer of copyright

Literary archivists can be asked to give advice when an unpublished document written in (for example) the 18th century – for which the testamentary history of copyright is unknown – is wanted for use as an illustration in a publication. In some cases copyright has not been specifically bequeathed, but has passed down several generations 'by default' rather than as willed property. Is Tim Padfield able to supply any information on whether there is any case law in this area? Such copyrights presumably go with the residue of the estate, but in our post-primogeniture times this will often involve them being split between many individuals, increasingly so as one generation succeeds another. What rights have the courts accorded individual descendants in these cases? Can a single one out of an unknown number of holders of a fraction of a copyright prevent publication of a text? If one of an unknown number of holders of a fraction of a copyright sues for damages, how is the fraction of the damages they are due calculated, when the number of co-holders is unknown?

3.1.7 Digitisation of literary manuscripts

Literary manuscripts are often prime candidates for digitisation; in fact digitisation has been promoted as a means of providing remote access to dispersed literary archives. However, the copyright barriers can be prohibitive. Can Tim Padfield verify that even if digitised images are uploaded in a way that prevents the downloading of copies, if copyright permission has not been obtained this still counts as a publication and breaches copyright? This presumably applies even to very early material if it has never previously been published, where copyright endures until 2039?

3.2 Case studies

3.2.1

A research library recently acquired the archive of a scholar and collector who was an expert on a particular nineteenth-century author, and was well-known for his literary detective work in tracking down sources for that author's fiction and locating primary sources. In addition to original manuscript material and correspondence, his

archive contains extensive quantities of photocopies: of manuscripts held in other repositories, and articles, chapters and cuttings from published books, journals and magazines. Most of these were supplied by libraries for private study, or copied by the researcher himself. Much of the photocopied material is taken from obscure sources, and provides a valuable ready-made resource for other scholars in the field; it also provides an important insight into his working methods. His archive would be fragmentary and much poorer without this material.

Should we have accepted the photocopied material at all, and can we make it available to researchers in the reading room? This also has relevance for cases (which are all too frequent in literary archives) where items have been 'cherry-picked' from an archive and sold off at auction (e.g. holograph manuscript drafts, or letters from highly significant individuals). Often such material can end up in private hands, and the photocopied versions provide the only means of access to the documents for researchers.

3.2.2

A research library holds an autograph manuscript of Oscar Wilde's early play *The Duchess of Padua*. The play was published in 1882 on the basis of this manuscript, which is marked up for the printer, and the published text is out of copyright, as Wilde died in 1900. However, the manuscript also includes many legible deletions which have never been published.

The Library wished to place images of pages from the manuscript on its website for promotional and informational purposes – including examples with interesting deletions. Our understanding was that these unpublished passages still had copyright protection and so we contacted the owner of Wilde's copyright, his grandson Merlin Holland, seeking permission to use the images. He responded that we should freely go ahead. So far as he was concerned, he said, the play and its manuscript were both out of copyright, despite the presence of unpublished material in the latter.

Which of us was correct?

4. The Data Protection Act and Literary Papers

4.1 Questions

4.1.1 Anonymisation and Research purposes/Section 33 of the DPA

There seems to be some confusion over the issue of anonymising the results of research carried out on records containing personal data – arising from misinterpretation of a reference in s33 of the Act, and possibly from a misunderstanding of the pro forma document 'Researcher undertaking concerning access to archives that would otherwise be closed' in the Society of Archivists' Code of Practice. The 'Legal Guidance' issued by the Information Commissioner's Office (p. 67) has also led to some confusion over this issue. It states:

‘As a matter of good practice, when processing for research, historical or statistical purposes, data controllers should always consider whether it is necessary to process personal data in order to achieve this purpose. Wherever possible, data controllers should only process data that has been stripped of all identifying features’.

This statement seems to be focusing on the kind of data gathered by a data controller for their own purposes, e.g. in medical or social surveys and questionnaires, rather than historical or literary research carried out using archives.

In some cases, the various pieces of guidance above have been interpreted as meaning that no documents containing personal data of living individuals may be made available to researchers unless the results of research on those data are anonymised. The stipulation to anonymise is obviously inappropriate in the case of literary research. Please can Susan Healy clarify that this is not the case, and that in fact the interpretation of the s33 exemption should essentially be as follows:

Personal data may be processed for research purposes, and kept indefinitely (exemption from Data Protection Principles 2 and 5) as long as the relevant conditions are met, i.e.

- the data are not processed to support measures or decisions with respect to individuals;
- the data are not processed in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject.

The other DP Principles should still be complied with.

4.1.2 Data subject access requests

Where material has been closed to readers under the DPA and an individual represented in that material makes a data subject access request, do we have the right to refuse this, e.g. if we believe that reading the material will cause substantial damage or distress to the data subject concerned? Literary archives are often personal papers; even if details of third parties are blanked out, it is likely that the data subject would be able to identify individuals who had made comments or committed actions that might cause damage or distress.

4.1.3 Structured filing systems

Should we assume that all personal data held by FOI public authorities (like HE institutions and national libraries) is covered by the Act, rather than just that which is organised in filing systems structured in such a way that individuals are readily identifiable by file titles? Much literary material takes the form of personal papers which are often highly unstructured, yet which often contain large quantities of sensitive material relating to living people.

4.1.4 ‘Substantial damage and distress’

Is there yet any case law or guidance which might help us to determine what constitutes 'substantial damage or distress'? Case studies 4.2.1-2 below provide some examples of how this can be problematic for literary archivists.

4.1.5 Remote enquiries

If we receive reprography requests from remote researchers (e.g. those based in countries which do not have data protection legislation in place) who wish to obtain copies of material which might relate to living individuals, should we also ask them to sign a data protection declaration as well as a standard copyright declaration before the copies are supplied?

4.1.6 The digital environment

Archives, including literary archives, are now increasingly being created digitally, and more institutions are developing the capacity to undertake digital preservation so they can acquire and make accessible archives in digital form. In the case of literary archives in the hard copy environment, DPA closures are usually made on an item-by-item basis. The sheer quantity of digital material is likely to make this impossible. Should we therefore assume that blanket closures must be applied to digital material which relates to living people? If so, are we contravening FOI by doing this? Should we include a standard clause in donation/deposit agreements for digital material, in which the donor/depositor explicitly stipulates that the material should remain closed for 100 years if it contains personal data of living individuals?

4.1.7 DPA and fictional works

Literary genres (poems, novels, short stories, dramas) produce unique forms of record types which are FICTIONS. How, if at all, does the DPA apply to works of fiction (or their prehistories in manuscript drafts)? How does the DPA apply to works of fiction (and their prehistories) where the names of "historical" individuals appear? Eg: a poet wrote a poem about the long and protracted death of the poet's (in his persona in the poem) wife from cancer. In one version of the poem in manuscript draft the protagonist is assisted to die. Whilst the poet and his wife are both dead, the drafts include two other characters whose names bear relation to "actual" people in the world. Since it is a poem and therefore a work of fiction there is no evidence whatsoever that the events conjured up actually happened. It is an imaginative construct exploring ideas about the ethical situation which characters find themselves in the poem. To extend the point, by way of another example, here is a passage from an interview which the great American poet Elizabeth Bishop gave to George Starbuck for *Ploughshares*, Spring 1977:

GS: I suppose Crusoe was a city kid. It's such fun, the accuracy with which you borrow flora and fauna for his little island. ("Crusoe in England," in *Geography III*).

EB: It's a mixture of several islands.

GS: And the deliberate anachronisms too — like the Wordsworth reference.

EB: *The New Yorker* sent the proof back and beside that line was the word "anachronism," and also at another place in the poem, I think. But I told them it was on purpose. But the snail shells, the blue snail shells, are true.

GS: Are there snails like that on — what was his island — Juan Fernandez?

EB: Perhaps — but the ones I've seen were in the Ten Thousand islands in Florida. Years ago I went on a canoe trip there and saw the blue snails. They were tree snails, and I still may have some. They were very frail and broke easily and they were all over everything. Fantastic.

See also case study 4.2.4 below, which demonstrates how the issue of literary genres also has particular implications within the context of fictional biography.

4.1.8 Closing documents under the DPA

When closing material we consider to be sensitive under the DPA, should we simply assume a closure period of 100 years from date of creation where information about birth date is not readily available?

4.2 Case studies

Case studies 4.2.1-2 deal with the issue of 'substantial damage and distress':

4.2.1

An openly Gay poet was crestfallen when I wrote to ask him if it would be ok to include a letter he'd written to his publisher in an exhibition. While he was happy for material relating to his sexuality to be made public, he objected to the use of this particular letter entirely on the grounds that it revealed he'd been poor when younger.

4.2.2

A prominent writer submitted a manuscript draft of his memoirs to his publisher which the publisher deemed to be potentially libellous. The writer changed the text to avoid the libel action; however, the original manuscript survives in the archive and its contents and libellous inferences also survive in references within the surrounding correspondence. The publisher pointed out that whilst the writer had potentially libelled a famous literary critic the critic would have been more offended (and therefore far more likely to take libel action) by that fact that the writer had insulted the critic's wife by describing her face as looking "like a red handbag". This kind of contextual information makes it difficult to determine what constitutes disproportionate damage and distress in DPA terms.

4.2.3

A famous avant garde novelist wrote a fictional biography which a publisher was concerned would be libellous. The writer changed some of the details of the text; however the original survives in manuscript draft in the archive, and the possible

libellous inferences also survive in references within the surrounding correspondence. This is a direct quotation from a letter sent by the publisher to a person “named” within the novel:

“When I was in my twenties I wrote a novel called *** which was partly based on childhood experiences. In it, the parents of the protagonist were made to seem monstrous. On the other hand, the actual locale of the novel, the character of the nurse and of other protagonists, was readily identifiable. My parents read the novel with horror since so much of it was “actual” but certain elements in it were painfully surprising to them. Indeed the book was not to “do the dirty” on them rather to draw out the nature of the relationship in which they were not involved. I understood their distress. On the other hand, the book was about something quite different and the “characterization” that they felt they had been subjected to had little to do with them, two individuals of whom I was very fond. I think the same must obtain in any book: there will always be recognisable figures and there will always be a sense of injustice when – as in Mann’s *Buddenbrooks* or Lawrence’s *Sons and Lovers*, the real is mingled with the fictional, the real is perceived from an alien perspective”.