Legal issues relating to the archiving of Internet resources in the UK, EU, USA and Australia

A study undertaken for the JISC and Wellcome Trust

Andrew Charlesworth
University of Bristol, Centre for IT and Law

Version 1.0 - 25 February 2003

JISC The Joint Information Systems Committee

The Wellcome Trust

Author Details

Andrew Charlesworth
Centre for IT and Law
Department of Law
University of Bristol
Wills Memorial Building
Queens Road
Bristol BS8 1RJ

Telephone: 0117 954 5355

Fax: 0117 925 1870

E-mail: a.j.charlesworth@bristol.ac.uk

Contents

Management summary and recommendations

Since its origins as a researcher's tool at CERN in the early 1990s, the World Wide Web has developed into an immense international complex of hyperlinked information. Some of the information available on the WWW simply mirrors that found in existing print publications. Much, however, is to be found nowhere else but (often temporarily) on the WWW. Some of that information, such as the webpages produced during and after the September 11 terrorist attacks, is of significant historical importance; other information, such as that found on medical websites, may be of long-term scientific value. The uniqueness of the information to be found on the medium, combined with the ephemerality of digital information, has resulted in a growing perception that there is a need for mechanisms to preserve at least some of that immense volume of information for the longer term.

The task of preserving web-based information is not, however, an easy one. Aside from the technical difficulties inherent in preserving transient digital resources, the legal environment in many countries is also often inhospitable to, or unappreciative of, the role of the would-be web archivist. If the most obvious legal stumbling-block is copyright law, hazards also lurk in the form of defamation law, content liability and data protection laws. Whilst these issues pose problems for the web archivist, these need not be insurmountable. Careful selection of resources, combined with an effective rights management policy, and processes for ensuring that controversial or potentially illegal material can be only selectively accessed, or can be removed if required, reduce significantly the likelihood of falling foul of the law or upsetting rightsholders. This paper examines the key legal issues in relation to the United Kingdom, and how potential risks to a UK based web archive might be minimised. It also surveys the approaches to web archiving taken in some other jurisdictions, including several EU countries, the US and Australia. The experiences in those countries suggest that:

- a legal framework for deposit or archiving of webpages is highly desirable to clarify legal issues and to protect the archivist (EU);

- a pragmatic approach to archiving can be successful, but will carry considerably heightened legal risks (US);

- in a jurisdiction where there is no legal framework for deposit or archiving of webpages, a licensing approach, while not able to cope with the breadth of material obtained by general harvesting, provides both an acceptable degree of legal risk, and permits the potential archiving of both 'shallow' and 'deep' web resources (Australia).

Audience and Purpose

This document is aimed primarily at archivists working in research institutions within the U.K. However many of the issues covered are of much broader scope than this and will be of relevance to archivists and web publishers both within and outside the U.K as well as archivists in other organisational settings. The purpose is to provide guidance on how to address the legal issues that will arise when creating a web archive from non-proprietary sources.

This document explains:

- Why the legal issues are important to archivists working with web resources;
- The need to develop a coherent approach to legal issues as part of webpage acquisition and preservation strategies;
- The latest legal developments of relevance to web archivists.

Legal Disclaimer

No part of this document constitutes formal legal advice, and it should not be used as a substitute for such. It contains interpretations of UK law and the law of other countries by the authors. No responsibility will be taken for the interpretation of this document by a third party. JISC and the Wellcome Trust strongly advise institutions and individuals to seek professional legal advice before taking any steps that might potentially breach UK law or compromise the intellectual property rights of others.

Report background

In March 2002, the Joint Information Systems Committee (JISC) and the Library of the Wellcome Trust invited proposals for an evaluation and feasibility study of Web archiving. The Wellcome Trust's interest in this subject is motivated by its interest in furthering medical research and the study of the history and public understanding of medicine. A proposal to extend the Wellcome Library's collecting activities to the Web has been endorsed by its Library Advisory Committee and the Medicine, Society and History Committee. The JISC's interest in Web archiving is prompted by its dual roles as a provider of Web-based services to the UK further education (FE) and higher education (HE) communities and as a funder of research and development projects. Both organisations are members of the Digital Preservation Coalition (DPC) and therefore committed to supporting collaboration to advance a common agenda in digital preservation.

In response to the JISC and Wellcome Trust's invitation, UKOLN undertook to produce a feasibility study into Web archiving. This aimed to provide the JISC and Wellcome Trust with:

- An analysis of existing Web archiving arrangements to determine to what extent they address the needs of the UK research and FE/HE communities. In particular this is focused on an evaluation of sites available through the Internet Archive's Wayback Machine, to see whether these would meet the needs of their current and future users.

- To provide recommendations on how the Wellcome Library and the JISC could begin to develop Web archiving initiatives to meet the needs of their constituent communities.

The feasibility study has resulted in the production of two separate reports:

- A general review of Web archiving issues and initiatives with recommendations for the JISC and Wellcome Trust by Michael Day of UKOLN. This outlines the urgent need for Web archiving initiatives and indicates the benefits these would have for the user communities of the JISC and Wellcome Trust. It includes an attempt to define to nature of the World Wide Web (and the UK part of it) and an introduction and evaluation of existing Web archiving initiatives. It ends with a short section on implementation.

- This study of legal issues by Andrew Charlesworth of the University of Bristol

Michael Day
UKOLN, University of Bath

1. Introduction

Web sites are an increasingly important part of [an] institution's digital assets and of [a] country's information and cultural heritage. (JISC – April 2002)

1.1. Preserving the Past

Even in the 'world of atoms'[1] the preservation of historical works can be a largely hit and miss affair. Despite the best efforts of librarians, archivists, curators and private collectors, many potentially important and influential works are lost to posterity due to oversight, neglect, decay and accidental or deliberate destruction. The value or influence of some works may simply not be understood at the time of their creation (or may be understood all too well), until historical events are re-evaluated by future generations of users, viewers and researchers. Considerable ingenuity may be required to collect, collate and preserve valuable collections of works - and even these collections are likely, by the very nature of their selection, to be but a partial record of their time.

The problems of such tangible collections are myriad, from corrosive ink in ancient manuscripts,[2] to rotting canvas and decaying pigments in paintings,[3] and 'vinegar syndrome' in triacetate film base[4] - these are but a few of the problems facing those seeking to preserve artefacts from the past.[5] How much simpler it might seem, to the untutored eye, to preserve modern digital artefacts - works that are easily copied at a high quality, stored in binary format on computer disk, diskettes, CD-ROM, and DVD. However, modern digital materials come with their own set of preservation problems. [6] The archivist or librarian must ensure that both the technical infrastructure and expertise necessary to read the materials remains available - the CAMiLEON project provides a good example of the difficulties inherent in maintaining such infrastructure and expertise.[7] Where compression or digital rights management technologies have been used, the means to unencrypt or unscramble the data is required. The use of techniques like hyperlinking means that maintaining the content and context of digital materials, such as webpages, is made more complex. The media on which digital material is held may also be subject to deterioration over time, requiring the transfer of data to new

---

1 Negroponte, N., *Being Digital* (London: Coronet, 1996)

2 The Iron Gall Ink Corrosion Website <http://www.knaw.nl/ecpa/ink/>

3 Kabbani, R.M. 'Conservation: A Collaboration Between Art and Science' *The Chemical Educator* 2 (1997): 1. <http://link.springer-ny.com/link/service/journals/00897/sbibs/s0002001/spapers/21rk897.pdf>.

4 Robley, L.P., 'Attack of the Vinegar Syndrome' *American Cinematographer*, June 1996. Reproduced at <http://www.capital.net/com/jaytp/VINEGAR.HTM>.

5 The American Institute for Conservation of Historic and Artistic Works, 'Basic guidelines for the care of special collections', 1999 <http://aic.stanford.edu/treasure/objects.html>.

6 Besser, H,. 'Digital Longevity' in Sitts, M., (ed.) *Handbook for Digital Projects: A Management Tool for Preservation and Access*, (Andover, Mass.: Northeast Document Conservation Center, 2000.)

7 The CAMiLEON Project <http://www.si.umich.edu/CAMILEON/>

storage media and possibly new formats.[8] As projects like CAMiLEON have shown, digital materials may in fact be less resilient than their tangible predecessors, and be more likely to vanish permanently within a relatively short space of time, unless particular efforts are made to preserve them. Copies of the original Domesday Book remain extant nearly a thousand years after its creation; copies of the BBC's 1980's Domesday Project, a pair of interactive videodiscs made by the BBC in London to celebrate the 900th anniversary of the original Domesday Book, and designed to capture a snapshot of British life in 1986, are now almost unusable:

> *[w]hile the 12" videodiscs are likely to remain readable for many years to come, the 1980s computers which read them and the BBC Micro software which interprets the digital data have a finite lifetime. With few working examples left, the 1986 Domesday Project is in danger of disappearing forever.*

A key advantage, in principle, is the fact that digital materials can often be copied rapidly, cheaply and perfectly. In practice, when considering the copying of digital materials for archival purposes, this advantage appears considerably circumscribed by two key problems, which can be described as technical and legal 'fencing'. Where the materials are produced or copied by the author or rightsholder with the intent of obtaining an economic return, extensive copying of the digital material might mean a significant reduction in that return.[9] The ease with which digital materials can be copied and the fidelity of the copies mean that those who create the materials (authors), or who have acquired property rights in them (rightsholders), risk losing control over their reproduction. This problem is, of course, not restricted to digital materials, and intellectual property law, especially the law of copyright, has long been used to provide a degree of protection for financial or intellectual investment in works of various kinds. Thus most national legal systems provide some form of legal 'fencing', usually by means of copyright law, to provide rightsholders with the power to control the extent to which users of digital material can make copies of it.

However, the degree of investment, skill and effort required to make an illegal copy of a digital work is often significantly lower than that which was previously required to make a copy of a tangible work. For example, prior to widespread digital music delivery, piracy of audio material on a scale that might seriously damage the interest of the rightsholders was limited mainly to large piracy operations that were relatively easily targeted under copyright law. In general, the scale of individual copying of audio material was limited, and because the recording technique used was normally analogue, the quality of the copy was poorer than the original. The arrival of digital music formats, combined with the connectivity of the Internet and the development of P2P technologies meant that significant economic damage could be caused to rightsholders by individuals making high quality audio material available to all and sundry for download on the Internet.

Thus, rightsholders in digital materials have increasingly begun to seek to retain control over their dissemination by restricting the ease of copying via technical means, or by acquiring additional legal controls over those acts that may be legitimately carried out with the materials by third parties. An example of 'technical fencing', or digital rights management (DRM), can be seen in the form of CDs that will play in hi-fi CD players, but not in CD players built into computers. The use of additional legal controls can be seen both in the changes to national copyright legislation and in the increasingly widespread use of contractual provisions to limit the copying and dissemination of digital works. Recent legislative developments, such as the

---

8 Ball, S., 'Magnetic and Digital Materials' Resource: The Council for Museums, Archives and Libraries <http://www.resource.gov.uk/information/advice/conserv14.asp>.

9 As exemplified by the difficulties that the music recording industry has with the making and exchanging of digital copies of music tracks over the Internet.

passage of the *Digital Millennium Copyright Act 1998*[10] in the US and the *Copyright Directive* (2001/29/EC)[11] in the EU combine the two approaches, by not just making it illegal to make a copy of a digital work without the rightsholder's permission, but also by making it illegal to remove or circumvent any technical controls placed on the work by the rightsholder to prevent copying, even copying that would be permissible under copyright law itself. In this arrangement, "[t]echnical protection measures facilitate the 'prevention' of unauthorised use of works, whereas copyright law is required to 'cure' infringements."[12] These new restrictions have inevitably affected the balance between the rights granted to the rightsholders and those granted to the general public under copyright law, as public rights found in national copyright legislation, such as fair use or fair dealing and library privileges are slowly eroded.

It is clear that in order to preserve digital materials in a useable form, librarians and archivists are likely to have to make copies of those works, whether as backups to the original works, or as replacements of the original works where it becomes necessary to migrate the work from one medium to another, or one format to another. While many national copyright laws appear to expressly permit the copying of existing tangible works to preserve or replace items in a permanent collection,[13] it is by no means clear whether this is necessarily the case for digital works, particularly where it is unclear if a work has actually been held in a 'permanent collection' - e.g. a library might now license access to an on-line database of periodicals for its users, instead of purchasing one or more paper periodicals for stack access.

The problem of preserving the Web is further compounded by the freedom to publish that it provides.[14] Prior to the Web, the ability to collate information in book, journal or pamphlet form, and then widely disseminate copies was effectively the exclusive province of increasingly monolithic firms of international publishers, and of government bodies and international organisations. This had two main implications for archivists. Firstly, the volume of information published was necessarily limited; secondly, the number of publishers with whom an archivist needed to negotiate was relatively small. The 'democratisation' of publishing brought about by the Web has led to an explosion in both the volume of information available, and the number of 'publishers' providing it. If volume alone had risen, the archivist could still have relied upon the established publishers to act as gatekeepers by exercising a measure of quality control, and ensuring legal compliance. However, increased volume combined with a multiplicity of publishers means that the would-be web archivist is now faced with significant and potentially costly feasibility/scalability problems.

As will be seen, without government intervention, for most archivists the choice will be to:

- create a selective archive by individual negotiation with rightowners or "publishers" which will be highly selective and cover a small percentage of what is available, but will carry lower legal risks and have relatively clear access rights.

---

10 *The Digital Millennium Copyright Act 1998* <http://www.loc.gov/copyright/legislation/dmca.pdf>

11 *The Copyright Directive* (2001/29/EC) - UK Implementation <http://www.patent.gov.uk/about/consultations/eccopyright/>

12 Anon., 'It was a dark and stormy night… E-Book distribution and copyright' <http://www.dcita.gov.au/Article/0,,0_1-2_1-4_15265,00.html>

13 See, for example, s.42 of the UK Copyright Design and Patents Act 1988 in conjunction with the relevant sections of The Copyright (Librarians and Archivists) (Copying of Copyright Material) Regulations 1989.

14 For a wealth of statistics and analysis on this topic, see Lyman, P. and Varian, H. R. "How Much Information", 2000. <http://www.sims.berkeley.edu/how-much-info>.

- create a more inclusive archive by means of automatic capture, which will result in the archiving of more sites and more information, but probably result in less quality control and also a correspondingly greater legal risk.

1.2. History in the Making

The World Wide Web ('the Web' or 'WWW') contains a tremendous amount of information contained in millions of webpages held on webservers distributed around the globe. Estimates vary considerably as to the numbers of webpages in the 'public Web' (sometimes referred to as the 'shallow Web') that is, those webpages which are open access as opposed to password protected, or part of subscription services (the 'private Web' or 'deep Web'), but it is clear that the corpus of this data is extremely volatile - undergoing constant daily change, whether by way of addition, amendment, or deletion. Anyone who has cited to webpages in their writings, or who has sought to use such citations, will be aware that much web content is ephemeral, appearing for short periods of time and then vanishing without trace.[15] While it might be fair to say that, for a significant percentage of Web content, such ephemerality is no great loss, it is equally true that the Web is potentially the source of a great deal of information of significant worth, be it historical, social or medical, and that the failure to adequately preserve at least some aspect of this immense potential archive would leave an unrecoverable gap in the historical record. The extent and implication of the possible loss has been compared to the early history of television, from which relatively little archival material remains.

The value of archiving, at the very least, selected portions of the Web was recognised at a relatively early stage of its development. The US Internet Archive (see below) has been collecting webpages since 1996, and currently archives over 10 billion pages in its web archive, including special collections dealing with the September 11 terrorist attacks and the US elections in 2000. Other organisations have taken a less expansive approach, for example, attempting to archive web pages in a specific domain, or on webservers in a specific geographical area. The Royal Library of Sweden's *Kulturarw*^{3} project (see below) aims to collect, preserve and make available Swedish documents from the Web, as part of the Royal Library's wider collection of printed publications collected since the 17th century.

1.3. Law and the Web Archivist

All Internet or Web archiving projects face the kind of technical difficulties outlined above when dealing with digital works. These difficulties have been described at length elsewhere.[16] This document will concentrate upon the legal problems that currently face those wishing to create a web archive. The obvious initial difficulty lies in the area of copyright - can the would-be archivist legally make archival copies of webpages, and if such archival webpages are made, what legitimate uses can be made of them during the term of copyright? Other difficulties arise not in terms of ownership of the content, but rather in terms of the legality of the content itself - to what extent is the archivist legally responsible for the illegal content stored on webpages in an archive? If an archived webpage contains defamatory material, or material which is potentially 'obscene' or 'indecent', or material which breaches the privacy rights of a third party, what liability might arise on behalf of the archivist?

---

15 One early estimate suggested the average lifetime of a web-based document was approximately 44 days. See Kahle, B., 'Preserving the Internet' *Scientific American*, March 1997. Quite how this time period was determined is unclear.

16 See, for example, Danish National Library Authority, *Preserving the present for the future: Proceedings of the Conference on Strategies for the Internet*, 18th - 19th June 2001, Copenhagen. <http://www.deflink.dk/upload/doc_filer/doc_alle/846_Trykt%20proceeding.pdf>

To some degree this will depend on the extent of the public access permitted to the web archive - an archive that is to all intents and purposes 'sealed' to public access for the foreseeable future runs a reduced legal risk with regard to its content - it cannot affect the rightsholders' economic rights in their works, and the lack of communication of the information contained within the archive means that issues of defamation and content liability are either removed or at least signally diminished. However, a web archive that does not permit public access loses much of its utility in the short to medium term, and it may be difficult to raise awareness of the archive's existence, and thus attract funding, unless some demonstrable immediate public benefit or gain can be shown.

The most effective solution to the general legal problems faced by archivists is likely to be national legislation. This may explicitly provide archivists with permission to make copies of works with the aim of preservation and archiving; provide for the legal deposit of works, both tangible and digital; and provide protections against prosecution for criminal offences, or civil suit merely for archiving certain types of work. However, national legislatures move very slowly, and the provision of legislation to make the task of archivists less onerous is rarely to the fore of legislators' minds. There is certainly neither clear international consensus on the legal status of archives, whether for tangible or digital works, nor a coherent international approach to harmonisation of legal rules affecting them. As will be outlined below, those organisations and individuals who have begun to archive parts of the Web have often done so either without a clear idea of the law which applies to their activities, or have taken the pragmatic, if not necessarily legally advisable, stance that they will continue to archive their particular areas until they run into specific problems with rightsholders or the authorities. Web archives see themselves offering a valuable service to future users, viewers and researchers, and often appear to simply hope that the beneficial aspects of their operations will persuade would-be litigants that they should be treated as a valued resource and not as copyright pirates, pornographers, or privacy breachers.

2. The United Kingdom

The UK does not, as yet, appear to have a significant web archive and certainly nothing on the scale of the US-based *Internet Archive* or the Swedish *Kulturarw*[3] or Danish *Netarchive.dk.* There thus appears to be little experience on which to draw when dealing with issues of UK law. Some information on dealing with preservation and archiving of digital works can be drawn from material available from the CEDARS and CAMiLEON projects, but there appears to be little or no widely available material aimed at the legal issues relating to web archiving.

2.1. Legal Issues

2.1.1. Copyright

In most jurisdictions, some degree of legal protection is provided to protect creative and innovative works against indiscriminate copying and use, by allowing individuals to claim rights in those works in a similar way to which they can claim rights in physical property. These rights are known as intellectual property rights (IPRs). Thus, the author of a manuscript can own a set of rights in his words that he can sell, lease, give away, or leave to his heirs just as he could sell, lease, give away, or leave to his heirs, a valuable piece of furniture. However, it is perhaps unwise to draw this analogy too far, as IPRs have several characteristics that are not easily equated to physical property, not least the ease with which they can be divided into smaller component rights, and the fact that a particular work may have more than one type of IPR attached to it.

There are a wide range of IPRs available, including some that are well known to the public, such as copyrights, patents and trademarks, and others that are less widely known, such as trade secrets, plant varieties, geographical indications and performers rights. IPRs often have to be applied for, the protection granted by them may be limited in scope to a particular country or trading area, and may vary in the degree of formality required according to national or international rules. While there is an increasing trend towards international harmonisation of IPRs, at present there are often wide disparities between different national and regional regimes.

Copyright is a property right vested in the owner of a protected work, and can be thought of as a bundle of economic rights and moral rights. It is a right that comes into being at the moment of creation of a work, and no formal procedure to register a copyright is required, or available, in the UK. Thus, under UK law, a copyright notice (© ABC 2002, All Rights Reserved") is not necessary, although many rightsholders use such notices to indicate their intention to defend their copyright in the case of infringements.

The basic framework of these rights is statutory, and contained in the *Copyright, Design and Patents Act 1988* (CDPA 1998), although the explanatory case law is of great importance. There is copyright protection for specific classes of works but not for ideas. Each type of work has a different status in law. Copyright law is a particularly complex subject, not least because copyright began life in the 1600s as a monopoly right for printers, and is now expected to cover material as diverse as artistic works and computer programs. The wide range of media that copyright law covers has led to a diversity of types and lengths of protection with which librarians and archivists would be advised to acquaint themselves, as each may require different strategies and considerations to obtain clearance for use.[17]

---

[17] See further, CEDARS, Cedars Guide To Intellectual Property Rights, 2002 <http://www.leeds.ac.uk/cedars/guideto/ipr/guidetoipr.pdf> and Padfield, T., *Copyright for Archivists and Users of Archive*s (Richmond: Public Records Office, 2001).

Copyright - Legal Deposit

Under the terms of the various UK Copyright Acts,[18] there are a number of copyright depositories which are entitled to claim a free copy of every print work published in the United Kingdom and Ireland, a process known as 'legal deposit'.[19] These depositories (the Bodleian Library, Oxford University, the University Library, Cambridge University, the National Library of Scotland, Trinity College, Dublin and the National Library of Wales) together employ an Agent, who claims on their behalf, and receives and distributes books and periodicals for them. The British Library is entitled to receive the material without claim and employs the Legal Deposit Office to manage this.[20] Neither the nature of the material which can be claimed under legal deposit nor the term 'published' are comprehensively defined within the Acts, but 'published' is generally taken to mean available to the public, and covers both priced and free material. Legal deposit does not presently apply to publication in microfilm, microfiche, CD-ROMs, databases or any of the newer non-book media.[21]

Thus, the legal deposit system in the UK does not currently extend to non-print or unpublished materials,[22] or to materials published outside the UK and Ireland. As such, the downloading and storage by one of the copyright depositories of material from a Web site, whether that site was based in the UK or elsewhere, would appear to be a straightforward infringement of copyright, in that such downloading and storage would inevitably involve the creation of unlicensed copies of the works that went to make up the webpage. In such circumstances, unless the agreement of the copyright owner was obtained in advance, web archiving in the UK without explicit permission from rightsholders would seem to place the budding archivist at risk of legal action.

For web archiving in the UK to be permissible via the legal deposit route, the law would have to be amended to include electronic materials, and the definition of 'publishing' more carefully defined so as to clearly cover works made publicly accessible on the WWW. In such circumstances the British Library, or other copyright depository, could potentially download and store electronic materials, such as webpages and make them available to patrons, although it is likely that any such law would need to be fairly restrictive regarding how many users could access the materials and what they could do with them. It should also be noted that the scope of legal deposit would still only cover materials published within the UK and Ireland, and that it might also be necessary to decide whether this included:

---

18 Currently the deposit privilege is based on s.15 of the Copyright Act of 1911, which remained unaltered by the Copyright Act of 1956, or the Copyright, Designs and Patents Act 1988. Although linked with copyright legislation for historical reasons, legal deposit is no longer connected in any way with the registration or ownership of copyright, or with copyright protection.

19 In practice, it appears that the copyright depositories are highly selective, even within those categories of print works that clearly fall under the Acts.

20 The Bodleian Library, General Principles of Collection Development and Access to Resources: Appendix 1 <http://www.bodley.ox.ac.uk/guides/bod/colldev.htm#APPENDIX 1>

21 See The British Library, Report of the Working Party on Legal Deposit, 2001. <http://www.bl.uk/about/policies/workreplegdep.html>

22 In this regard the UK trails some other countries, such as Canada, France, Germany, Iran, Italy, Japan, Sweden and the United States, which already include electronic publications in their legal deposit scheme, making off-line electronic publications subject to legal deposit in the form of the depositing of a physical item or a publication in a fixed format.

- materials published on a webserver based physically within the UK or Ireland,
- materials published on a webserver with a UK or Irish-based domain registration, regardless of physical location
- materials published on a webserver and publicly accessible to UK and Irish citizens

In any event, amendment of UK copyright legislation to permit legal deposit of any of the 3 categories listed above would still be subject to the caveat that the UK government can only legislate on issues within its jurisdiction. The third category of materials, those published on a webserver and publicly accessible to UK and Irish citizens, would therefore be a controversial category to grant a sweeping legal deposit power over, as they would inevitably include materials created, stored and subject to IPR regimes outside the UK jurisdiction, and would thus be theoretically outside the scope of UK legislation.

Copyright - Library and Archive Copying

UK copyright legislation also makes explicit provision for both library and archive copying for preservation and replacement. The *Copyright, Designs and Patents Act 1988* s.42 (see Appendix) permits libraries and archives to make a copy from any item in their permanent collection for preservation and replacement. A prescribed library for the purpose of making a copy to replace a copy of a work under s42 includes all libraries in the UK.[23] A prescribed archive for the purpose of making a copy to replace a copy of a work under s42 includes all archives in the UK.[24]

However, with regard to digital archiving, especially web archiving, the legislation as currently worded is not terribly helpful. Merely having access to a webpage does not make that webpage part of a library or archive's permanent collection, and thus the rights provided to libraries and archives under s.42 CDPA to make copies of works without the permission of the rightsholder, would appear inapplicable to archiving of web pages.

Copyright - Licensing

In the absence of a statutory right to archive webpages, either by means of legal deposit, or under other library and archive privileges, it would appear that the only way that the webpages in the UK can be archived in conformity with copyright law is for the would-be archivist to endeavour to obtain the necessary permissions to copy the works in particular webpages from the relevant rightholders. Within a relatively restricted domain, such as a University website, this might be feasible, as many of the works within such a website would most likely already belong to the institution, if created by the University's employees in the course of their employment, or if created by contractors with whom an assignment of copyright had been agreed.

On the other hand, the wider the archivist seeks to cast her net, the more difficult the task of obtaining the relevant permissions becomes, as it becomes difficult to effectively track and record who exactly is the rightsholder for particular material. To some extent this could be ameliorated by establishing intermediary rights management, e.g. by requesting that website owners take responsibility for ensuring that there is sufficient legal metadata on their webpages to identify both the ownership, and the level of permission to copy, of the content on their websites. However, this too is likely to prove unwieldy in practice on anything more than the most compact archives, and would impose a burden on website owners that they may

---

23 Statutory Instrument 1989 No. 1212 The Copyright (Librarians and Archivists) (Copying of Copyright Material) Regulations 1989 reg. 3(2).

24 SI 1989 No. 1212 reg. 3(4)..

not care to carry, particularly if they are indifferent as to whether their website, or collection of webpages, is preserved for posterity.

Copyright - Opt-out

A further possibility might be for a web archivist to archive their chosen websites and webpages without requesting prior permission from rightholders, but instead supplying rightholders with the opportunity to opt-out of having their webpages archived. This might take the form of an:

- *A priori* opt-out - here rightholders who do not wish all, or part, of their website or collection of webpages to be archived, indicate, via some agreed code in their webpages, that this is the case. An example of such a system can be seen in the form of the Web Robots Exclusion Protocol,[25] and the Web Robots META tag.[26]

- *A posteori* opt-out - here websites and webpages are archived without the prior permission of the rightsholder, but a clear mechanism is provided by the archivist to allow rightholders to request the removal of their work or works.

However, while these methods may seem to indicate a solution to some of the problems, neither of them allows the archivist to avoid the reach of copyright law. The *a priori* opt-out requires the rightsholder to make an indication as to the copyability of their works that rightholders are not obliged to make under copyright law. Under UK copyright law there need be no indication that a work is copyright, as a work meeting the necessary criteria for protection automatically receives it. Failure to indicate a preference cannot therefore be taken to override the rightholder's protection under copyright law. Equally, the *a posteori* opt-out requires action on the part of the rightholder, to seek out the removal mechanism and use it, which copyright law does not require, and worse, the unlawful copying has already taken place before the rightholder has had a chance to object. While a combination of both mechanisms might assuage the majority of rightholders, or at least cause them to forego the potentially expensive route of legal action in the light of a less expensive option, there is nothing to stop a determined rightholder ignoring both *a priori* and *a posteori* opt-outs and still being able to bring a successful suit against the archivist.

2.1.2. Defamation

In most, if not all, jurisdictions, the fundamental basis of defamation liability is the publication of untrue information, that liability will be based on the extent of the damage to the reputation of the person referred to in that information, and that a person's reputation cannot be damaged unless the information is disseminated to other people than the author. English law[27] imposes liability regardless of whether the publisher of a statement knew or ought to have known it was defamatory[28] Unlike English law, Scots law[29] provides that the defamatory statement need

---

25 Which indicates to web robots which parts of a site should not be visited, by means of a specially formatted file in http://.../robots.txt.

26 Which allows a webpage author to indicate if a page may or may not be indexed, or analysed for links, through the use of a special HTML META tag

27 See further Price, D., *Defamation: Law, Procedure and Practice* (London: Sweet & Maxwell, 1997)

28 *Hulton & Co. v. Jones* [1910] AC 20.

29 See further Norrie, K. *Defamation and Related Actions in Scots Law* (London: Butterworths 1995)

only be communicated to the pursuer for an action to lie and justify an award of at least nominal damages.[30]

Defamation - Libel

Libel consists of a defamatory statement or representation in permanent form, anything which is temporary and audible only is slander. Statements in books, articles, newspapers and letters are libels, as are statements in e-mails and webpages. For a statement to be libellous, it must be:

- defamatory as opposed to vituperative/abusive
- refer to the plaintiff in such a fashion that the plaintiff can be clearly identified
- made known to others or 'published'. Publication in English libel law terms takes place when information is disseminated to other people than the author and the plaintiff.

The key legislation in this area is the *Defamation Act 1996*, which was designed to simplify and modernise the law of defamation, in particular with regard to determining who could be sued in a given action (see Appendix). s.1 of the Act appears, in part, to have been designed to provide a specific defence for Internet Service Providers (ISPs) and other Internet Intermediaries (IIs) who transfer data without exercising any editorial function, although the effect of the section depends heavily on an "all reasonable care" test. For those with an authorial or editorial role in publishing on the Internet, the law of libel applies just as it does to the print medium.

The application of the law to an Internet web archive suggests the following points:

- the display of false information damaging to the reputation of the person referred to in that information, on a public webpage, will be considered by the courts to be published, and thus libellous;
- the author of the statement on the webpage may be sued for damages, unless they did not intend their statement to be published at all;
- if the statement is published on a website which is edited (or moderated), i.e. some other person than the author has control over the content of the statement or the decision to publish it, that "editor" may be sued for damages;
- if the statement is published on a website by a commercial publisher defined in the Act as "a person whose business is issuing material to the public, or a section of the public" - s.1(2) - there is no requirement of payment by the public - that publisher may be sued for damages;
- if the person 'publishing' the statement on the website is not the author, editor or publisher because they do not fit the respective definitions in s.1(1)(a), or because they are merely involved in "processing, making copies of, distributing or selling any electronic medium in or on which the statement is recorded, or in operating or providing any equipment, system or service by means of which the statement is retrieved, copied, distributed or made available in electronic form" or acting "as the operator of or provider of access to a communications system by means of which the statement is transmitted, or made available, by a person over whom he has no effective control" - s1(3) they may not be sued for damages UNLESS

---

30 *Mackay v McCankie* (1883) 10 R 537.

- they failed to take reasonable care in relation to its publication, or knew, or had reason to believe, that what they did caused or contributed to the publication of a defamatory statement - s.(1)(b) and s.1(1)(c), in which case they too can be sued.

The *Godfrey v. Demon Internet Ltd* case[31] provides a graphic example of how, despite the exemption in s.1, ISPs and IIs can fall foul of the law. In this case, a posting in the USA was made to an Internet newsgroup "soc.culture.thai" which Demon Internet carried and stored. The message was forged such that it appeared to come from the plaintiff. The plaintiff notified Demon Internet that the posting was a forgery and requested them to remove the posting from their Usenet news server as it was defamatory of him. Demon Internet failed to remove the message, although they could have done so, and it remained available on their news server until its expiry some 10 days later. While Demon Internet would appear to fall within the definition of a 'publisher' under s 1(2) of the Act, they sought to argue that they were exempted from liability by s.1(3) and s.1(1)(a). The judge agreed with this but noted that they were also subject to s.1(1)(b) and 1(1)(c) of the Defamation Act 1996 and that following the plaintiff's notification they were unable to argue that they had taken reasonable care with regard to the publication, and did not know and had no reason to believe that they were causing or contributing to a defamatory statement. In the words of the Lord Chancellor' Department

> *The defence of innocent dissemination has never provided an absolute immunity for distributors, however mechanical their contribution. It does not protect those who knew that the material they were handling was defamatory, or who ought to have known of its nature.*[32]

Additionally, the *Loutchansky v Times Newspapers Ltd and Others (No 2)* case[33] demonstrates that the current law of defamation in the UK may pose specific problems for archive providers. It is an established principle of English libel law that each individual publication of a libel gives rise to a separate cause of action, subject to its own limitation period.[34] s.4A of the *Limitation Act 1980* provides that

> *no action for libel or slander, slander of title, slander of goods or other malicious falsehood shall be brought after the expiration of one year from the date on which the cause of action accrued.*

In the *Loutchansky* case, Loutchansky sued over articles appearing in The Times on 8 September 1999 and 14 October 1999 which accused him of certain criminal activities. Each article was posted and archived on The Times' website. Following complaints by Loutchansky that the articles were still available via the website, a qualification was added to the online version of the first article on 23 December 2000. The warning alerted readers to the fact that the article was 'subject to High Court libel litigation' and cautioned that it 'should not be reproduced or relied on without reference to Times Newspapers Legal Department.' The Times sought in court to claim the benefit of common law qualified privilege.[35]

> *Qualified privilege can protect anyone who makes a defamatory statement in the performance of a legal, moral or social duty, to a person who has a*

---

31 QBD, [1999] 4 *All ER* 342.

32 Lord Chancellor's Department, *Reforming Defamation Law and Procedure: Consultation on Draft Bill*, July 1995, paragraph 2.4.

33 CA, [2002] 1 *All ER* 652.

34 See *Duke of Brunswick v Harmer* [1849] 14 QB 185

35 See *Reynolds v Times Newspapers Ltd* [2001] 2 AC 127.

> *corresponding duty or interest in receiving it. The potential for applying this formula to statements published in newspapers might be thought to be obvious; news reporters see themselves as under a duty to report events of which it is in the public's interest to be informed.* [36]

However, the defence afforded by common law qualified privilege is dependant upon the publisher demonstrating a duty to publish potentially defamatory words to the world at large, and the Court held that in determining whether this was the case, the standard to be applied was that of responsible journalism. In the case of the on-line articles, the Court felt that the failure of *The Times* to attach any qualification to them on its website, over the period of a year, and despite the ongoing litigation, could not be described as responsible journalism and thus for the articles on the website no qualified privilege defence could be claimed.

During the litigation, *The Times* also argued that the limitation period in relation to the online version of the articles had begun to run as soon as they were first posted on the website, and that as Loutchansky commenced defamation proceedings in relation to the online versions of the articles on 9 December 2000, this period had expired prior to the commencement of those proceedings. This was rejected by the Court, which noted that it was a well-established principle of English defamation law that each individual publication of a libel gives rise to a separate cause of action, subject to its own limitation period. *The Times* argued that this rule was in conflict with Article 10 of the European Convention on Human Rights, because it has a 'chilling effect upon the freedom of expression which goes beyond what is necessary and proportionate in a democratic society for the protection of the reputation of others.'

However, the Court of Appeal ruled that:

> *… we accept that the maintenance of archives whether in hard copy or on the Internet has a social utility but consider that the maintenance of archives is a comparatively insignificant aspect of freedom of expression. … nor do we believe that the law of defamation need inhibit the responsible maintenance of archives … where it is known that archive material is or may be defamatory, the attachment of an appropriate notice warning against treating it as the truth will normally remove any sting from the material.*

The effects of the *Loutchansky* decision are that:

- for the purposes of s.4A of the *Limitation Act 1980*, on-line archives are in effect being continuously republished. As such, defamatory material accessible via the Internet could be the subject of legal action in England long after the original date of publication as re-publication lays the publishers open to legal action every new day that the defamatory statement appears.

- in order to minimize the risk of ongoing liability for defamatory material stored in an online archive, publishers should remove or disable access to that material immediately after the commencement of defamation proceedings, or attach a warning to the material noting that it is the subject of defamation proceedings and that the truth of the material is contested.

Defamation - The Electronic Commerce (EC Directive) Regulations 2002

The *Electronic Commerce (EC Directive) Regulations 2002* (see Appendix) were laid before Parliament on 31 July 2002 and largely came into force on 21 August 2002. The regulations

---

36 Legal500.com, Recent developments in common law qualified privilege <http://www.legal500.com/devs/uk/en/uken_051.htm>

are intended, amongst other things, to transpose articles 12, 13 and 14 of the EU *Electronic Commerce Directive* concerning the liability of Internet intermediaries for carrying, caching or hosting information provided by others, and will potentially provide statutory defences for Internet intermediaries in respect of defamatory material which they carry, cache or host, but which they did not create - regulations 17, 18 and 19. However, regulation 22 clearly provides that those defences in regulations 18 and 19 for intermediaries who cache or host defamatory Internet material which they did not create will ordinarily be defeated where the intermediaries are put on notice, even by e-mail, of the existence of the offending material.

The government has said it is prepared to consider including in the future additional regulations providing protection from liability for other categories of intermediaries, such as providers of hyperlinks, location tools and content aggregation, but has rejected calls for the inclusion of a regulation transposing article 15 of the Directive on Electronic Commerce which would prohibit the imposition of a general obligation on intermediaries to monitor the information they transmit or store, or a general obligation actively to seek facts or circumstances indicating illegal activity.

The *Electronic Commerce (EC Directive) Regulations 2002* would not appear to change the legal situation as regards web archives, as a person or organisation providing a web archive is not a 'mere conduit', is not engaging in 'caching' within the meaning of the Regulations, and would seem to fall outside the definition of 'hosting'.

Defamation - Notice and Takedown

It seems clear therefore that the web archivist must pay careful attention to the nature of her archiving operations. There are various possibilities available:

- If the archivist simply archives all the data on all the webpages visited by her web robots without exercising any editorial function whatsoever she may not be considered to an author or an editor - there is unlikely to be liability for defamation should one of the archived webpages contain a defamatory statement.

- If the archivist makes decisions about the webpages that are archived, she may be seen to be exercising an editorial function. This might even be the case where the decision to archive or reject a page is carried out by a web robot, on the basis of certain programmed choices made by the archivist - there might be liability for defamation should one of the archived webpages contain a defamatory statement.

- If the archivist simply archives all the data on all the webpages visited by her web robots without exercising any editorial function BUT provides public access to the resulting web archive, she may still be considered a publisher for the purpose of s.1(1)(a) under the definitions in s.1(2), and thus liable for publishing a libel should one of the archived webpages contain a defamatory statement - there might be liability for defamation should one of the archived webpages contain a defamatory statement.

- If the archivist simply archives all the data on all the webpages visited by her web robots without exercising any editorial function BUT provides public access to the resulting web archive, she may still be considered a publisher for the purpose of s.1(1)(a) under the definitions in s.1(2) BUT be exempted by virtue of reliance on s.1(3) - there is unlikely to be liability for defamation should one of the archived webpages contain a defamatory statement.

- If the archivist simply archives all the data on all the webpages visited by her web robots without exercising any editorial function BUT provides public access to the resulting web archive, she may be exempted from liability as a publisher by virtue of reliance on s.1(3), BUT ONLY if she has additionally taken reasonable care as regards the content of the web archive, and she has made provision for dealing with situations where she is put on notice by a third party or parties that material she is carrying may be defamatory. Failure to remove defamatory data from the publicly accessible archive, or to attach a warning to

the material noting that it is the subject of defamation proceedings, and that the truth of the material is contested, might lead to liability for defamation should one of the archived webpages contain a defamatory statement.

It would appear therefore that the web archivist would be wise to have a procedure in place for accepting notice from individuals complaining they have been defamed, e.g. a clearly identifiable person responsible for handling such complaints, the clear provision of contact address and other contact details for that person, and an effective mechanism for handling any complaint that should arise, either by way of immediate posting of a warning on the information complained off, or more likely by its immediate removal from public access until such time as there is no longer reason to believe that the material is defamatory. Such a process could be part of a wider 'notice and take down' procedure for other types of contentious material in the archive, such as material that infringes copyrights, and material containing illegal content.

Defamation - Offer to make amends

In the event that a defamation action is threatened as a result of the publicly accessible archiving of a webpage containing defamatory information, the archivist may wish to attempt to avoid litigation by:

- Issuing an apology, either verbally or in writing - the person who has been defamed may be prepared to accept an apology rather than undertaking expensive legal action. Such an apology might also involve the removal of the offending material and an undertaking not to publish it again.

- Under the UK *Defamation Act 1996* ss2-4, if an apology is not accepted then an offer to make amends may be made to the person defamed. This can be made either before or after the complainant has started court action. A valid offer to make amends must be made in writing; be expressly made under s.2 of the *Defamation Act 1996*; and state whether it is a qualified offer, that is, whether it relates to only part of the alleged defamation. It is an offer to publish a suitable correction and a sufficient apology and to pay the claimant compensation and costs. If the offer to make amends is not accepted by the claimant, then it will be a defence to defamation proceedings unless the claimant can prove that the defendant knew or had reason to believe that the statement complained of: referred to the claimant or was likely to be understood as referring to him; and was both false and defamatory.

Defamation - Jurisdiction

The Internet is an international medium, and a web archive accessible via the Internet and not domain limited, or otherwise restricted as regards access, risks exposing itself to multi-jurisdictional liability. The fact that a message or webpage may be accessible from, or downloaded in, another country may be enough for its courts find jurisdiction and to accept a legal claim there - collecting damages awarded in another state is of course, another matter.

2.1.3. Content Liability

There are various other types of potential content liability that may cause the web-archivist problems. Not the least of these is the issue of material of objectionable content, whether pornographic, violent or otherwise distasteful to some part of the archive's audience. In the UK, the primary pieces of legislation dealing with this type of material are the *Obscene Publications Acts* of 1959 and 1964[37] and the *Protection of Children Act 1978* (as amended by

---

[37] In Scotland, where the Obscene Publications Act does not apply, the Civic Government (Scotland) Act 1982 makes it an offence to publish obscene material and prosecution is the responsibility of the Procurator Fiscal Service. The Obscene

s.84-87 of the *Criminal Justice and Public Order Act* 1994)[38] The *Telecommunications Act 1984* also contains some relevant provisions in s.43.

Content Liability - Obscene Publications

The *Obscene Publications Act 1959*, s.1(1) states that

> *an article shall be deemed to be obscene if its effect . . . is, if taken as a whole, such as to* tend to deprave and corrupt *persons who are likely . . . to read, see or hear the matter contained or embodied in it.'*

This test bears considerable similarity to that in an 1868 court decision, *R. v. Hicklin*,[39] where the judge stated that whether an article was obscene or not depended upon

> *whether the tendency of the matter … is to deprave and corrupt those whose minds are open to such immoral influences and into whose hands a publication of this sort may fall.* [40]

It is clear that this legal definition of obscene has rather more specific meaning than would normally be attributed to the definition of obscene in non-legal usage. It is important also to remember that while the depiction of sexual acts in pictorial or textual form is the most obvious form of potentially obscene material, the caselaw demonstrates that, for example, action may also be taken against aural presentations such as music albums,[41] pamphlets advocating the use of drugs,[42] and material showing scenes of violence.[43]

The key issues to consider when assessing particular material are:

- The possibility of the relevant material being seen as likely to deprave and corrupt.
- Could an observer come to the conclusion that some of those who viewed the material might be depraved and corrupted by it?
- The likely audience for the material, as this will form part of the assessment of its tendency to deprave and corrupt.

When deciding whether material is obscene, an important determining factor is the consideration of whom its likely audience is going to be. This is because some potential

---

Publications Act 1959 also does not extend to Northern Ireland. Obscene material, including video works, is generally dealt with under the common law offence of publishing an obscene libel.

38 See s.172 (8) for those parts of the Act applicable to Scotland.

39 (1868) L.R. 3 Q.B. 360, 371.

40 Quoted in Heins, M. *Indecency: The Ongoing American Debate Over Sex, Children, Free Speech, and Dirty Words* The Andy Warhol Foundation for the Visual Arts Paper Series on the Arts, Culture and Society Paper Number 7; <http://www.warholfoundation.org/paperseries/article7.htm> [visited 15/08/02].

41 'Singled out for abuse', *Independent*, August 8, 1991, 17; 'Niggaz court win marks changing attitude', *Guardian*, November 8, 1991.

42 *Calder v. Powell* [1965] 1 QB 509, *R v Skirving* [1985] QB 819.

43 *DPP v. A & B Chewing Gum* [1968] 1 QB 119.

audiences are regarded as being more susceptible to being depraved and corrupted than others. Children are seen as an audience that is especially vulnerable in this respect. Thus, material available in a forum or media that is open to children will be always be subject to stricter regulation than material that is not. Material on the Internet is obtainable in relatively uncontrolled circumstances, and thus the definition of what is likely to deprave and corrupt those likely to have access to the Internet will be accordingly low.

If an article is obscene, it is an offence to publish it or to have it for publication for gain. The *Obscene Publications Act 1959*, s.1(3) as amended by the *Criminal Justice and Public Order Act 1994*,[44] defines a publisher as one who in relation to obscene material:

> (a) distributes, circulates, sells, lets on hire, gives or lends it, or who offers for sale or for letting on hire, or.
>
> (b) in the case of an article containing or embodying matter to be looked at or a record, shows, plays or projects it, or, where the matter is data stored electronically, transmits that data

The transfer of obscene material either manually by use of computer disks or other storage media, or electronically from one computer to another via a network or the Internet clearly falls under section 1(3). The *Obscene Publications Act 1964*, section 1(2) makes it an offence to have an obscene article in ownership, possession or control with a view to publishing it for gain.

Obscene material placed on a webserver will be caught even when an individual simply makes the data available to be transferred or downloaded electronically by others so that they can access the materials and copy them. This was demonstrated in the case of *R v Arnolds, R v Fellows*[45]. On appeal from their initial conviction, the defendants argued that the act of placing material on an Internet site could not be regarded as a form of distribution or publication. The Court of Appeal, however, held that while the legislation required some activity on the part of the 'publisher', this seemed to be amply provided by the fact that one of the appellants had taken,

> *whatever steps were necessary not merely to store the data on his computer but also to make it available world wide to other computers via the Internet. He corresponded by e-mail with those who sought to have access to it and he imposed certain conditions before they were permitted to do so.*

The two main defences to obscenity charges contained in the *Obscene Publications Act 1959* are innocent publication and publication in the public good. Innocent publications means that the person who published the material in question did not known that it was obscene and had no reasonable cause to believe that its publication would result in liability under the Act s.2(5)). In the Internet context, it can be seen that while a provider of facilities or Internet Service Provider is unaware that obscene material is being put onto the Internet via their system they cannot be liable. However, if they are put on notice that this is occurring, they will have to take action to bring the activity to a halt. Failure to take such action would leave them at significant risk of prosecution. An example of this has been the activities of the police in putting Internet Service Providers on notice of Usenet newsgroups that contain potentially obscene material.[46] This provides great impetus to the Internet Service Providers to drop such newsgroups, as the notice would make it virtually impossible to run a successful defence of innocent publication. In contrast to providers who host webpages or newsgroups, those

---

44 s.168 and Schedule 9, para. 3.

45 [1997] 2 All ER 548.

46 *The Independent*, 20 December 1995.

providers who simply provide a connection to the Internet are unlikely to be able, even if they wanted to, to be in a position to accurately assess the nature even a fraction of the data that their systems carry. They are thus unlikely to incur liability, even if some of their users use their systems as a conduit to access or distribute pornography, as there can be no actual knowledge of the material carried.

The defence of public good is found in s.4 of the *Obscene Publications Act 1959* which states that:

> *publication of the article in question is justified as being for the public good on the ground that it is in the interests of science, literature, art or learning, or of other objects of general concern*

The defence does not mean that the article is not obscene, but rather that the obscene elements are outweighed by one of the interests listed. As may be gleaned from the discussion of the definition of pornography above, much may be read into the context in which the purportedly 'obscene' material is to be found. Indeed, the first case to arise under the legislation, in 1961, concerned D.H. Lawrence's book *Lady Chatterley's Lover.* Undoubtedly, some of the passages of the book were rather explicit for the period, but taken as a whole, the book's clear literary merits, which were defended by a number of experts, helped ensure its acquittal. It has been argued that, in some cases, the concept of literary merit has been rather liberally construed, for example, the book *Inside Linda Lovelace*, about the porn actress who starred in *Deep Throat*, was cleared on similar grounds in 1976.

A key problem with the Obscene Publications Acts is that the only certain way to test whether or not material is obscene, or if it is obscene whether it serves the public good, is via the courts. A good example of the difficulties this creates was an incident in June 1998 when British police seized a book, *Mapplethorpe*, from the stock of the library at the University of Central England in Birmingham. The book contained photographs of homosexual activity and bondage scenes taken by the internationally renowned photographer and artist Robert Mapplethorpe. Despite the fact that the book was widely acknowledged as serious artistic work, the police told the University that its contents might contravene the *Obscene Publications Act 1959*. The book came to the attention of the police when a student at the University's Institute of Art and Design took photographs of prints contained in the book to a local chemist for developing and the chemist forwarded them to the police. Ironically, the student had taken the photographs to include them in a thesis entitled 'Fine Art versus Pornography.' It seems that the police, at least, had little doubt as to their interpretation. This is a clear example of a work which in the eyes of a significant element in society (the police) is clearly obscene, and in the eyes of others (the University of Central England) a work of artistic merit. The uncertainty that this generates tends to have a 'chilling' effect on the nature and scope of material that is created, published, and distributed, in the UK, as publishers and other distributors are less willing to publish controversial material.

Content Liability - Indecent Publications

The relevant parts of the amended *Protection of Children Act 1978* (PCA) deal with photographic representations of children under 16 (or persons who appear to be under 16). The Act makes it an offence to take, make, permit to be taken, distribute, show, possess intending to distribute or show, or publish, indecent photographs or pseudo-photographs of children. The Act defines 'distribution' very broadly. It is not necessary for actual possession of the material to pass from one person to another, the material merely has to be exposed or offered for acquisition. The PCA also criminalises advertisements which suggest that the advertiser distributes or shows indecent photographs of children, or intends to do so. The legislative amendments made by the *Criminal Justice Act 1988* further criminalise the mere possession of such photographs or pseudo-photographs.

s.84(4) of the *Criminal Justice and Public Order Act 1994* (CJPOA) inserted a subsection (b) to s.7(4) of the 1978 Act stating that 'photograph' shall include:

> *data stored on a computer disc or by other electronic means which is capable of conversion into a photograph.*

While this definition of photograph covers digital representations of physical photographs (thus gif and jpeg image files, downloaded from FTP sites, embedded in webpages, or compiled from Usenet messages, will be treated as photographs), it was not considered sufficiently broad. s.84 of the CJPOA added the concept of the pseudo-photograph:

> *Pseudo-photograph" means an image, whether made by computer-graphics or otherwise howsoever, which appears to be a photograph.*

Thus a pseudo-photograph means any image which is capable of being resolved into an image which appears to be a photograph and, if the image appears to show a child, then the image is to be treated as if that of a child. This means that there is no need for a child to have been used in the creation of the image, indeed the Act covers an indecent image which may not be based on any living subject. The pseudo–photograph amendments deal with situations where, for instance, morphing software is used to create images which look as if they are of children from images of adults. Given the increasing difficulty of detecting faked photographs, and the tendency of defendants to argue that individuals in seized images were not in fact children, this change seems logical. Some have argued that the purpose of the PCA was to prevent harm coming to actual children, and if no children are used in the making of pseudo-photographs, such photographs whether indecent or not should remain outside the law. Others counter that paedophiles have been known to use indecent photographs to persuade children that unlawful sexual activity is acceptable behaviour, and thus children may be harmed by the existence of such material.

Unlike obscenity, the term 'indecency' is not defined in either the PCA, or any other statute in which it occurs. When one examines statutes which refer to indecency, such as those which prohibit, the import of indecent materials (see the *Customs Consolidation Act 1876*, the *Customs and Excise Management Act 1979*), or sending such materials through the post (the Post Office Act 1953), or their public display (the *Indecent Displays (Control) Act 1981*) it appears that 'indecency' relates to material that is considered 'shocking and disgusting', but less 'shocking and disgusting' than material which is considered obscene. In practice, the test for indecency remains just as subjective, and thus just as difficult to pin down, as that for obscenity. In essence, the test would seem to be whether the item in question offends current standards of propriety, or to put it in the American phraseology, whether it offends contemporary community standards.[47] Given that community standards of adult behaviour tend to be rather higher where children are involved, an image involving a naked adult which might be perfectly acceptable could well be treated as indecent if a child or pseudo-child image were to be portrayed in a similar manner.

The provisions discussed above have clear relevance to activities on the Internet. It would seem to follow from the *Arnold* case mentioned above, that placing of indecent pictures of children on a webserver will almost inevitably mean that they will be distributed; when such pictures are held on a computer they can be plausibly said to be in someone's possession; a link to a web site may be considered an advertisement; and a e-mail offering such pictures in digital or paper form certainly would.

A person or company charged under the PCA with distributing, showing, or possessing intending to show or distribute, has two potential defences, the first being that the person or company charged did not see the image and that they had no knowledge or suspicion that the image was indecent, and the second that there was a legitimate reason for possessing or distributing the image e.g. for academic research.

---

47 See *United States v. Thomas* 74 F.3d 701 (6th Cir. 1996).

It is also an offence to possess an indecent image of a child or indecent child-like image. The defences available are to be found in the amended version of s.160 of the 1988 Act. These are similar to those contained in the PCA, but include what might be termed an 'unsolicited indecent material' defence:

> (1) It is an offence for a person to have any indecent photograph or pseudo-photograph of a child in his possession.
>
> (2) Where a person is charged with an offence under ss(1) above, it shall be a defence for him to prove -
>
> (a) that he had a legitimate reason for having the photograph or pseudo-photograph in his possession; or
>
> (b) that he had not himself seen the photograph or pseudo-photograph and did not know, nor had any cause to suspect, it to be indecent; or
>
> (c) that the photograph or pseudo-photograph was sent to him without any prior request made by him or on his behalf and that he did not keep it for an unreasonable time.

With regard to the computerised making or possession of indecent photographs of children, the UK courts held in *R v. Bowden* that the intentional downloading and/or printing out of computer data of indecent images of children from the Internet constituted the 'making' of an indecent photograph and was thus an offence under s1(1)(a) of the Protection of Children Act 1978.[48] With regard to the unintentional storage of computer data of indecent images of children in a computer cache the court in *Atkins v DPP* held that this did not automatically constitute 'making', nor did their possession in a computer cache necessarily mean an offence had been committed under s160 Criminal Justice Act 1988, as the defendant, in such circumstances, must be shown to have known he had the photographs in his possession, or to know he once had them.[49]

In *R v Smith and Jayson*,[50] Smith had received an indecent photograph as an email attachment, and Jayson had browsed an indecent pseudo-photograph on the Internet. In both cases, their browser software automatically saved the images to a temporary Internet cache on their computers. With regard to Smith, the court held that no offence of "making" or "being in possession" of an indecent pseudo-photograph was committed simply by opening an email attachment where the recipient was unaware that it contained or was likely to contain an indecent image, noting than in *Atkins* it was held that the Act did not create an absolute offence encompassing the unintentional making of copies. However, when Smith's opening of the e-mail attachment was considered in the light of the evidence relating to his other activities, the court did not believe him to be unaware of the nature of the attachment. In regard to Jayson, he argued that his act of viewing the indecent pseudo-photograph did not constitute the necessary intent to 'make' a photograph or pseudo-photograph. The court, however, held that the act of voluntarily downloading an indecent image from the Internet to a computer screen was an act of making a photograph or pseudo-photograph, as the intent required was 'a deliberate and intentional act with the knowledge that the image was or was likely to be an indecent photograph or pseudo-photograph of a child.' Thus, Jayson did not have to intend to store the image with a view to future retrieval in order to meet the intent requirement for 'making'.

---

48 [2000] 2 All ER 418

49 [2000] 2 All ER 425, 436.

50 7 March 2002 (CA)

Content Liability - Defending Preservation

Whilst some would argue that material potentially falling within the scope of obscene material should not be archived,[51] such material does form part of the historical record, and additionally, given changing cultural and moral standards over time, some material that is considered obscene today may eventually be seen in a rather different light - sometimes even as art or literature. Even that obscene material which is unlikely to ever be considered 'artistic' may be of use to the future ethnographer. For example, the increasing availability of 'hardcore' pornographic material on the Internet has influenced the degree to which such material is available in print form - adult magazines have pushed the boundaries of what was previously permissible, in order to retain their audience, and governments and regulators have increasingly turned a blind eye to print material which is now freely available on-line. Additionally, as far as the UK is concerned, the legal availability of 'hardcore' pornography in other Member States of the EU has led to a gradual relaxation of national rules on import and dissemination of such material in print and video form.[52] A hypothetical future researcher might well be able to trace the effect of these trends via a study of UK webpages of the period.[53]

However, such considerations notwithstanding, a web archive which contains material (pornographic or otherwise) that could potentially 'deprave and corrupt' some element of those using it will have to consider carefully its access and use policies. The key issue in UK law is the target audience - the wider the audience the more stringent the controls will need to be to ensure that the obscenity test is not breached. If archiving of websites for a web archive is largely carried out by automatic processes, the archivist will be faced with a number of potential options:

- Limiting collection to a known or 'trusted' set of webpages. This will work with subject specific web archives, where the archivist has already largely determined what will be archived and from where. However, the larger and less selective the archival process, the higher the probability that potentially obscene material will be collected accidentally.

- Ensuring the collection software does not collect certain types of material. This is a difficult task, not least because the software tools currently available lack the discrimination to make the necessary determinations with sufficient accuracy. For example, although some filter software companies have produced software that can filter photographs by the amount of 'flesh' coloration in the picture, with the aim being to block pornographic material, this remains a very hit and miss technology.[54]

---

51 See, for example, the controversy surrounding the National Library of Australia's decision to include pornographic material from the web in the PANDORA digital archive. < http://news.bbc.co.uk/1/hi/world/asia-pacific/2221489.stm>

52 For example, see Travis, A., *Bound and Gagged* (London: Profile, 2000) for a discussion of the changing nature of the R18 classification used by the British Board of Film Classification to classify adult film & video. <http://www.bbfc.co.uk/website/Customers.nsf/Guidelines/GuidelinesTheCategories R18?OpenDocument>

53 However, this is a fairly simplistic example, as there are other significant influences on the likely content of UK pornographic webpages, for example, the standard usage policies of the average UK ISP. At present, a vanishingly small percentage of pornographic websites in the UK are hosted by UK ISPs, or held on UK-based servers.

54 Wilson, M. 'Artificially intelligent strategies for filtering offensive images on the Internet', April 29, 2001 <http://www.cs.indiana.edu/~marawils/writing/aiporn.html>

- Limiting access to all or part of the archive by minors. While this would decrease the chance of material likely to 'deprave and corrupt' reaching those that the law would seek to protect, it may be difficult to organise in practice - age verification on the Internet can be an inexact science.

- Providing a 'take-down' procedure. As with copyright-breaching and defamatory material, the web archivist could have a procedure in place for accepting notice from individuals about material in the website that might be considered to be obscene or indecent. This approach would, however, require the archivist to take relatively rapid action upon notification, for while she is unaware that obscene material is being held in her archive she cannot be liable for it, but once she is notified that defence is lost.

- Arguing the defence of public good. In the case of a web archive, it may well be possible to argue that the archiving process 'is in the interests of science, literature, art or learning, or of other objects of general concern' and thus that the harm of any obscene elements in the archive will be outweighed by the public good/interest in having an accurate record of the particular webspace archived. On the other hand this would probably carry more weight if backed by one or more of the other measures listed above.

With regard to potentially indecent material, there will only rarely be any justification for retaining such material in a web archive - it is possible that some medical photographs of children might be acceptable within a medical web archive, but might be unacceptable if provided for wider circulation - although the availability for archiving of such pictures on the public web would seem unlikely. In circumstances where potentially indecent material is collected accidentally, and its existence becomes known to the archivist, the material should be removed from the web archive immediately and the appropriate authorities notified. Destruction of the material should, however, be left to the authorities, as immediate destruction by the archivist might hinder criminal investigations against the original supplier. In practice, it is unlikely that indecent, as opposed to obscene, material will be found on the public web, as it appears that much of this material is supplied through private websites and FTP servers that will be inaccessible to cataloguing and harvesting software.

2.1.4. Data Protection

Over the last 40 or so years, the increasing computerisation of data relating to individual citizens, whether by government or by private enterprise, has been viewed with increasing alarm by those who see such computerisation as potentially leading to considerable breaches of an individual's right to privacy.[55] Technical advances in the use of such data, by means of techniques such as data matching and data mining,[56] have allowed seemingly disparate sources of personal information to be aggregated and examined in ways that those who initially provided the data probably never envisaged. This tension between the social utility of freely accessible personal information (and Western society depends ever more heavily on the free flow of such information for the operation of elements as disparate as the social security and banking systems) and the perception that this may lead to unwarranted or unfair invasion of an individual's informational privacy has led many states to pass legislation restricting the collection, storage, and use of personal data.[57] A web archive will inevitably contain large

---

55 See for example, Garfinkel, S., *Database Nation* (Sebastopol: O'Reilly, 2000); Jennings, C. & Fena, L., *The Hundredth Window: Protecting Your Privacy and Security in the Age of the Internet* (New York: Free Press, 2000); Hunter, R., *World without Secrets: Business, Crime and Privacy in the Age of Ubiquitous Computing* (John Wiley & Sons, 2002).

56 Delmater, R. & Hancock, M., *Data Mining Explained* (Digital Press, 2001).

57 Rotenberg, M., *The Privacy Law Sourcebook 2001* (Electronic Privacy Information Center, 2001).

amounts of personal data - webpages often contain names, addresses, work and home telephone numbers, archives of posts to message boards, live chat forums, Usenet and e-mail mailing lists, and a myriad other pieces of information relating to personal facts and figures. The collection and storage of these pieces of data in a web archive may result in personal data becoming, and indeed remaining, available for search and retrieval when at the time of dissemination the individual concerned neither knew that this might occur, nor would have wanted such an outcome.

Data Protection - The effect on the Web

In the UK, the *Data Protection Act 1998*, which implements the EU *Data Protection Directive 1995* into UK law, and the considerable secondary legislation pursuant to that Act, has to be considered. Unfortunately, the Act, like the Directive, already looks dated in some respects with regard to the use of modern data technologies. Neither, for example, applies well to the Web, as any webmaster faced with the fact that publication of material on the public web almost certainly means publication to countries outside the EEA without 'adequate' levels of DP protection, can attest. In the UK, at least, such problems have been addressed with a considerable degree of pragmatism by the Office of the Information Commissioner largely in terms of risk/benefit assessment, i.e. is there a benefit to the data controller vs. the likelihood of substantial damage or substantial distress to any individual. This has not necessarily been the case in other EU countries.

The issue of the impact of the DPA 1998 on archives in general has been addressed in guidelines by both the Public Records Office[58] and the Society of Archivists.[59] While these documents provide a useful background to the application of the Act to archiving, neither of them deals explicitly with large or small scale web archiving. Given that the drafters of the legislation almost certainly were not thinking of the Web when drawing up its provisions, it is hardly surprising that those interpreting the legislation are now wary of entering this arena. The documents do provide some guidance that might with help with regard to the application of the Act to web archiving, but much of this is less than reassuring to the would be web archiver.

It is clear that anyone who holds information about readily identifiable living individuals has to comply with data protection law in managing that information. A web archive will inevitably contain such data. The nature of a web archive will also almost certainly require the web archivist, as a data controller, to notify the Information Commissioner of the archivist's intention to process personal data and to keep this notification up-to-date. The type of processing carried out by a web archive is unlikely to allow for a general description of the processing of personal data under the headings set out in the Commissioner's *Notification Handbook*, so it may be that a web archive would fall under the following special purpose description which has been approved for archives by the Commissioner:

> *Records selected for permanent preservation as archives, with a view to their use in historical or other research.*

Although this description covers the archives of private sector bodies, for example, those of businesses or private research institutes, or of individuals, it is difficult to ascertain, however, the extent to which it might be held to cover archives that are not built from the records owned

---

58 PRO, *Data Protection Act: A Guide for Records Managers and Archivists*, 2000. <http://www.pro.gov.uk/recordsmanagement/dp/dpguide.pdf>

59 Public Record Office, Society of Archivists & Records Management Society, *Code of Practice for Archivists and Records Managers under Section 51(4) of the data Protection Act 1998*. Version 2, 20 April 2002 <http://www.archives.org.uk/publications/soacodev2.doc>

by such businesses or private research institutes, or individuals, but rather harvested from other sources by software agents.

Should a suitable notification heading be found, it would appear difficult to apply the core principles of the Data Protection Act to a web archive, even given the exemptions provided to archives generally, for the Act assumes a much greater knowledge and control of the personal data in an archive than the web archivist may be able to provide. Certainly, strictly applying the DP principles to a web archive of the size of the US *Internet Archive*, containing as it does over 100 terabytes of data and with a growth rate of 12 terabytes a month, would seem to be impossible.

The Society of Archivists Guidelines suggest that:

> *As a general rule archives received by an archives repository can fall into any of three categories: […] Gifts, legacies or purchases, the common factor being that ownership of the archives passes to the archives repository or its parent organisation. […] Deposits from external sources, whereby custody passes to the archives repository but ownership remains with the depositor. […] Transfers from within the organisation, which may be a public authority or a private sector body such as a business.*

Web archives do not appear to necessarily fall within any of these categories - where the webpages are harvested from all or a selection of the public web by software agents, the information containing in them is not gifted or purchased, it is probably not deposited in any formal sense of the word, as the external sources may not have given any *a priori* permissions, and it is not derived from within an organisation.

Data Protection - The Data Protection Principles

Acquisition and processing of personal data (Principles 1 and 2)

The Society of Archivists Guidelines state that:

> *4.2.2 Processing for the purposes of archival preservation is undertaken by reference to the "research exemptions" set out in s.33 of the Act. […] Provided that the "relevant conditions" are observed, namely:*
>
> - *That the data are not processed to support measures or decisions with respect to particular individuals, and*
> - *That 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*
>
> *personal data may be stored indefinitely as archives for research purposes. The section exempts the data controller from the requirement to comply with Principles 2 and 5 but the other Principles must be observed. The data may be disclosed to third parties for research purposes or to the data subject without this exemption from compliance being lost.*
>
> *4.2.3 All archives repositories acquiring personal data and wishing to further process them must be able to show that there is a "fair" and "lawful" basis for doing so, in accordance with Principle 1. […]*
>
> *4.2.4 Archivists processing sensitive personal data who are unable to comply with any of the conditions specified in Schedule 3 may benefit from SI 2000 No. 417* Data Protection (Processing of Sensitive Personal Data) Order 2000, *which sets out additional circumstances in which it is lawful to process sensitive personal data. Paragraph 9 of the Order makes lawful processing which:*

> *(a) is in the substantial public interest;*
>
> *(b) is necessary for "research purposes" (which expression shall have the same meaning as in section 33 of the Act);*
>
> *(c) does not support measures or decisions with respect to any particular data subject otherwise than with the explicit consent of that data subject; and*
>
> *(d) does not cause, nor is likely to cause, substantial damage or distress to the data subject or any other person. […]*

> *4.2.6 […] archivists will generally not be expected to inform the subjects of data they (further) process for research purposes because to do so would involve disproportionate effort. The unfairness of not so informing data subjects is minimal where records are either to be kept closed for a long period or to be used only for research which will be anonymised. However, it would be unfair not to inform a particularly famous individual of the processing of his data if he himself was not the donor or depositor and the data are being dealt with in a special way, e.g. published, which is not happening to the rest of the archive.*

With respect to large web archives created by harvesting webpages from the public web, this advice would appear to be difficult to apply in practice, especially if those webpages are then made available to the general public in a searchable form. It is debatable to what extent the wholesale archiving of webpages might be provably "in the substantial public interest", although more specific collections, such as those relating to political, cultural or medical issues might be more easily justified. Determining in advance whether archival processing does or does not "support measures or decisions with respect to any particular data subject otherwise than with the explicit consent of that data subject" or might "cause, or be likely to cause, substantial damage or distress to the data subject or any other person" is unlikely to be feasible.

Maintaining accuracy of personal data (Principle 4)

The Society of Archivists Guidelines state that:

> *4.7.1 […] personal data preserved as archives are not expected to be kept "up-to-date" in the same way as data still subject to operational use. Archives are concerned with historical integrity rather than current accuracy. It seems likely that in the event of legal proceedings brought by a data subject over inaccuracy, the court would order data to be supplemented by a statement of the true facts. Archivists should be able to rely on the use of supplementary statements or certificates to make the rectification without damaging archival integrity.*

This approach would appear to be potentially feasible with regard to a web archive - a data subject concerned that an archived webpage contained misleading or inaccurate personal data could be provided with a mechanism by which he could automatically attach an amendment to the page, or could be provided with a contact person who would deal with such issues. The guidelines suggest further that, in the event that data is held for archival purposes, the Information Commissioner would be less likely to press for the right of data subjects to block, erase or order the destruction of personal data they believe to be false (as opposed to amendment by supplementary statements or certificates) to be applied.

Data subject access to personal data (Principle 6)

The Society of Archivists Guidelines state that:

> *4.8.1 Archivists who are data controllers (or joint data controllers) will be responsible for providing data subject access to personal data covered by the Act. […]*
>
> *4.8.2 Although archivists may find they have no legal obligation to respond to a data subject access request, for example when the records concerned are held for archival preservation purposes only and are not open for research, it is nonetheless good practice to provide the data as a matter of policy, especially if the rights and entitlements of individuals are at stake. […]*

In an open access web archive, data subjects would be able to search for their own personal data and if necessary make a request for amendment by supplementary statements or certificates, or in extreme cases for blocking, erasure or destruction of inaccurate personal data. In a closed archive the issue is moot. However, in an archive open only to researchers, it appears that the archivist might find herself required to undertake searches of the archive on behalf of a data subject.

Security of personal data (Principle 7)

The Society of Archivists Guidelines state that:

> *4.4.1 All newly received archives (manual and electronic) should be checked to ascertain whether they include personal data covered by the Act, for example a series of case files about named living individuals. Appropriate storage and access conditions should be applied to these archives*

In a large web archive, unless the process of checking suggested can be automated to a high degree, this advice is likely to be unworkable. In smaller subject specific web archives this might be possible, but would still be time and resource intensive.

Transfer of personal data outside the EEA (Principle 8)

A web archive like the US *Internet Archive* which is accessible and searchable on the web inevitably involves the transfer of personal data to any nation in the world from which the *Internet Archive* can be reached. From a UK point of view, accessibility of a UK based web archive from the EEA would be covered by the assumption that EEA nations have legislation in compliance with the EU Data Protection Directive - all other countries would be subject to assessment as to whether their data protection laws were 'adequate'. Where a country's law is inadequate, personal data should not be exported from the EU to that country, unless some other mechanism for protection of the data subjects' rights is provided, e.g. sectoral protection or contractual protection. Very few countries are currently deemed to have 'adequate' data protection regimes.

Data Protection - Opt-out

The admittedly brief analysis above suggests that ensuring compliance with the UK data protection regime is going to be difficult for a web archive. Some of the webpages that are harvested will almost certainly contain personal data. Data subjects may be unaware that their personal data is on the web, or unaware that it may be collected and archived. The personal data placed on the web may have been placed there by third parties without permission and may be accidentally or deliberately (and perhaps maliciously) inaccurate. It may be a mix of 'ordinary' or 'sensitive' personal data, and without careful human checking, its precise nature may be impossible to verify. While the law provides some protection for archivists, that protection does not seem to have been designed with web archiving in mind, being aimed at more traditional forms of archiving.

Without clear guidance from the Information Commissioner's Office, and on the face of the legislation, it appears that the web archivist may run a significant risk of having her processing deemed to be in breach of the Act. In such circumstances, the Information Commissioner might issue an enforcement notice to prevent further processing. Enforcement notices require a data controller to take steps to ensure compliance with the data protection principles.[60] This may require her to stop processing any personal data, or certain types of personal data, or to stop processing personal data, for a particular purpose, or in a particular way.[61] It may also require her to rectify, block, erase, or destroy inaccurate data and any other data held by her that appears to be based on the inaccurate data,[62] and, if practicable, to notify third parties to whom the data have been disclosed that they have been rectified, blocked, erased, or destroyed.[63] An enforcement notice must state which principle or principles have been contravened, and the Commissioner's reasons for her conclusion.[64] Individuals who are concerned that their personal data are, or may be, being processed in a manner that contravenes the Act may also request the Commissioner to investigate.[65] If the Commissioner considers that the person making a request has a legitimate and timely concern, she may make an assessment of the processing involved, to see whether it is in compliance with the Act.

It might be possible to alleviate some of the potential problems by providing the mechanism suggested above, by which a data subject concerned that an archived webpage contained misleading or inaccurate personal could either automatically attach an amendment to the archived webpage, or be provided with a contact person who would deal with such issues. However, the provision of such a mechanism, whilst it might be a useful way to defuse some data subject complaints, would not, and could not, offer a complete solution to the data protection issues.

2.2. Existing Archives and Policies

While there has been a least one pilot practical study on web archiving, run by the British Library, this appears to currently be a very small scale operation covering only 100 UK-based websites (although there are apparently plans to scale up the archiving to 10 000 websites), and presently there appears to have been little feedback on the legal implications of the study.

2.3. Future Developments

In three of the key legal areas discussed above there is a degree of re-evaluation and change in train that may have some effect on the ease with which web archiving can be undertaken.

The *European Copyright Directive*[66] entered into force on 22 June 2001. Article 13 requires the Directive to be transposed into the national laws of EU Member States before 22 December 2002. Implementing the Directive will involve a major overhaul of UK copyright

---

60 s.40, DPA, (1998).

61 s.40(1)(b), DPA, (1998).

62 s.40(3), DPA, (1998).

63 s.40(5), DPA, (1998).

64 s.40(6), DPA, (1998).

65 s.42, DPA, (1998).

66 Community Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society <http://europa.eu.int/eur-lex/pri/en/oj/dat/2001/l_167/l_16720010622en00100019.pdf>

law, and there is currently a UK government consultation, run via the UK Patent Office, to discuss the precise way that UK law in this area should be amended in order to comply with the Directive. However, this consultation period will only remain open until 31 October 2002. The changes thus far suggested by the government are, at best, unlikely to make web archiving any easier under UK copyright law, and at worse may place further obstacles to the web archivist. The UK government has also made favourable noises about extending the scope of legal deposit to digital works, which would probably include websites, but as yet no firm legislative action has taken place.

The UK Law Commission is in the process of re-examining the law of defamation in the UK, including the implications of the *Godfrey v. Demon Internet* and *Loutchansky v Times Newspapers Ltd and Others (No 2)* cases. Preliminary advice to the government on whether legislative change is required is likely to be finalised in the next 2-3 months.

The EU Commission is currently evaluating the impact of the EU Data Protection Directive with the aim of deciding whether aspects of the existing law (and national implementing legislation) may need to be altered in the light of present and future issues. Here too, initial decisions as to areas of change are likely to be made in the next 2-3 months.

3. The European Union

While the European Union has had a significant influence over the development of certain aspects of the law relevant to web archiving, most notably in the areas of data protection law and copyright law, there is no uniform approach to archiving or to legal deposit across the Member States. Even in the spheres of data protection law and copyright law, where the EU Commission's aim has been to harmonise the laws of the Member States in order to prevent obstacles to the free movement of goods and services within the EU internal market, it has tended to propose legislation in the form of Directives. These provide a broad indication of the aims that the Commission wants to achieve but, by their very nature, permit the Member States significant leeway in how those aims are to be achieved in national implementing legislation.

The end result of this has been that there are often wide divergences between the supposedly harmonised Member State systems. This is particularly clear in the area of data protection law, where the Commission is currently considering further measures to harmonise Member State laws following the various implementations of the 1995 Directive. It remains to be seen how coherent an EU-wide system of copyright law will emerge in the wake of the recent Copyright Directive, but the combination of a mix of common and civil law traditions, and the generally piecemeal approach to IPR legislation in the EU, does not seem likely to provide a clear and comprehensive system in the near future.

The issues of content control and defamation show even less uniformity across the Member States, as might be expected given the fragmented nature of the EU on matters of acceptable types of cultural and social discourse. Some Member States operate rigorous regimes of censorship over depictions of sexual activity, whilst others, like the Netherlands, prefer a rather more *laissez faire* approach to their citizens' proclivities in this area. Even in those areas of moral judgement where some degree of consensus might reasonably be expected, such as the undesirability of child pornography, the extent of that consensus does not appear to extend to the uniform interpretation of subject matter, uniform definition of offences, or uniformity of punishment, across the EU.

3.1. Legal Issues

Given the above discussion, it is clear that, short of providing a synopsis of the legal categories explored in the previous sections with regard to the UK, for each of the EU Member States (a task well beyond the scope of this report), a comprehensive report will be impossible to deliver. One can state with some degree of certainty that data protection laws based on the Data Protection Directive 1995 and enforced to a greater or lesser degree by both national data protection commissioners and national courts exist in all the Member States. As such, a web archivist in any of the EU Member States will be subject to broadly similar rules, albeit with widely differing interpretations and degrees of enforcement.

Similarly, one can reasonably infer that as all the Member States have copyright laws, based largely on the basis of the Berne Convention and related WIPO treaties, and harmonised to a certain degree by various EU copyright legislation, wholesale copying of webpages without the permission of the rightholders of those pages, by a web archivist in any of the EU Member States, will be open to some degree of civil and/or criminal sanction.

In the area of illegal content, it might be fairly claimed that the UK represents the most censorious end of the obscenity/pornography scale in the EU, but beyond that it is difficult to generalise. As regards indecent material, defined for this purpose as 'real photographs of actual minors engaged in sexual activity', it is probable that all the Member States would consider it to be undesirable, and thus probably not an appropriate type of information to be archived and made available for public viewing even 'in the public interest'. Should the

Council of Europe's somewhat controversial Cybercrime Convention[67] ever be ratified by enough states to bring it into force, a clearer definition of child pornography would likely emerge, but at present, this seems some way from actuality.[68]

With regard to defamation law, most EU countries deal with defamation under civil law. As noted above, it is possible to state with some certainty that in most, if not all, jurisdictions, the fundamental basis of defamation liability is the publication of untrue information, that liability will be based on the extent of the damage to the reputation of the person referred to in that information, and that a person's reputation cannot be damaged unless the information is disseminated to other people than the author. Once one ventures beyond these basic principles, national defamation laws rapidly diverge, for example, under Finnish law a distinction is made between intentional and negligent defamation, in the UK there is no such distinction.

3.2. Existing Archives and Policies

There are at least 3 web archives[69] currently in existence across the EU. Each has a limited remit related to websites that clearly or plausibly fall within national jurisdiction and in all cases the archives are carrying out their work within the formal framework of a legal deposit scheme for digital works.

3.2.1. Denmark - Netarchive.dk and the Royal Library

Netarchive.dk was a one-year project investigating strategies for collecting and archiving Danish Internet materials, running from August 1, 2001 to July 31, 2002, and was carried out by the Royal Library, Copenhagen, The State and University Library, Aarhus, and the Centre for Internet Research at the University of Aarhus.[70] The project was aided by changes in the Danish legislation on legal deposit in 1997[71] which brought Internet material within the scope of works which could be collected and archived. The legislation defined a "work" as being a delimited quantity of information that is considered a final[72] and independent[73] unit, "published" as being "when, with the consent of the author, copies of the work have been placed on sale or otherwise distributed to the public" or when "notice is given to the public

---

67 CoE, Convention on Cybercrime <http://conventions.coe.int/Treaty/en/Treaties/Html/185.htm>

68 The Treaty requires 5 ratifications (of which at least 3 must be by Member States of the Council of Europe) to enter into force - at present it has 2, Albania and Croatia <http://conventions.coe.int/Treaty/EN/CadreListeTraites.htm>

69 The 3 national web archives described in this section are those for which reasonable amounts of recently updated information are available on the Web. Finland (Helsinki University Library - Project Eva), Germany (Deutsche Bibliothek), Austria (AOLA) and the UK (British Library - Domain.uk) have also run pilot harvesting schemes - all these schemes have been hampered by the lack of a clear legal framework for web archiving, notably in the area of legal deposit. None appear to have taken a broader view of the potential legal pitfalls.

70 netarchive.dk <http://www.netarchive.dk/index-en.htm>

71 See The Danish Law of Legal Deposit (undated, page source suggests May 2002) <http://www.kb.dk/kb/dept/nbo/da/pligtafl/information-en.htm>

72 'Final' is interpreted to mean 'not continually updated'.

73 'Independent' is interpreted to mean as 'not part of a major work'.

that copies of the work are being produced and will be distributed to order, or that the work is available from a database from which the user can retrieve a copy", and stated that works covered by these definitions could be subject to legal deposit "regardless of medium".

Prior to the project, legal deposit of Internet publications was already underway, with deposits being collected by the Royal Library in Copenhagen. The depositor for a web publication is the person in charge of the technical completion of the digital copy. The depositor does not actually deposit the work but notifies the Royal Library of publication through an on-line registration form which has 3 versions:

- For monographic works with metadata
- For monographic works without metadata
- For periodicals

The Royal Library checks the document and, if it is covered by the law, downloads the document and places it on the archival server. The depositor gets two receipts (by e-mail): one after notification and the other after successful download. This approach is hampered by the fact that, unlike conventional publishers, many prospective web depositors are unaware of the existence of legal deposit, despite mail campaigns and newspaper advertising.

Additionally, it appears that Danish law requires that for an Internet document to be subject to legal deposit it must be 'static' (e.g. completed or only occasionally updated monographs and periodicals) rather than 'dynamic' (e.g. databases and homepages). This has meant that both the Royal Library and the netarchive.dk project could only archive static documents.[74] As static documents make up only a small percentage of the Danish web,[75] this is something of a stumbling block to the effective proposed archiving of the .dk domain, and makes it difficult to create a fully automatic system via which all relevant web material can be harvested and registered.

The Royal Library material is archived in the form that it is received and without modification. When provided to users via the Library's display system, it is supplied through a database in such a way that all URLs are corrected to references within the archive instead of to active documents on the web. Due to copyright legislation, the Library is not allowed to give access over the web to deposited digital works and the archived net publications can only be viewed at the reading rooms in the legal deposit libraries where print-outs for personal use are allowed.

The netarchive.dk website itself is not particularly informative about the project outcomes, and only one of the expected 4 project reports has been translated into English, although three in Danish have now been mounted. Both the Royal Library and the netarchive.dk materials appear to concentrate on copyright issues to the exclusion of other legal issues.

3.2.2. Sweden - Kulturarw[3]

The Kulturarw[3] project is run by Sweden's Royal Library and has been in operation since 1996. Kulturarw[3]'s approach is premised on that of the private and non-profit Internet Archive Foundation in San Francisco (see below), and the project aims to preserve as much as possible

---

74 Henriksen, B. 'Danish Legal Deposit: Experience & the Need for Adjustments' <http://www.deflink.dk/upload/doc_filer/doc_alle/1023_BNH.doc>

75 The Royal Library estimated that "There are currently some 300 000 registered Danish domains. In the [Royal Library] archive are net publications from less than 1 000 domains" The Danish Law of Legal Deposit *op.cit.* n. 71.

of the 'Swedish web'.[76] The project decided not to operate an archive limited to specific types of website or web document, because:

- of the difficulty in establishing exactly which material would be of value to future researchers and which would not.
- of the potentially high cost of a selectivity exercise in terms of staff time and costs
- the decreasing cost of media for data storage makes such large scale archiving feasible and cost effective.

The archive saves everything found within the ccTLD '.se', as well as Swedish owned web sites among other TLDs such as '.org', '.net' and '.nu'. Those additional web sites are selected manually, if physically located in Sweden, or if of Swedish interest.[77] The archive currently only saves material from the public Internet and thus does not archive webpages requiring passwords. There is no selection as regards the type of document acquired, i.e. all picture, sound and other file types are collected.

When visiting websites to harvest data Kulturarw^3 obeys site-based instructions/limitations on what may be acquired and indexed, i.e. robots.txt files and robots metadata. Kulturarw^3 argues, however, that such instructions/limitations are usually devised with an indexing robot in mind. As such pictures and short-lived material are often blocked for access because pictures cannot be indexed and short-lived pages will have disappeared before they are indexed and loaded into the database. Kulturarw^3, however, would wish to archive such material, and it is suggested that are many cases where Kulturarw^3 would like to ignore such site-based instructions/limitations. Presently Kulturarw^3 chooses to obey them as the legal framework for its activity remains unclear, and to ignore site-based instructions/limitations would be a clear breach of 'netiquette'.[78]

The Swedish government issued a special decree in May 2002, with regard to the work done by the Royal Library in acquiring, preserving and making accessible everything found on the Swedish Internet. Prior to this, the Royal Library had collected web materials on the premise that the existing legal framework permitted such collection, but had refused public access to the material. The decree authorizes the Royal Library to both collect material from Swedish web sites on the Internet and also to allow public access to it within library premises.[79] The legal discussion surrounding Kulturarw^3, as with Netarchive.dk appears to be exclusively focused on legal deposit and copyright.

---

76 Kulturarw^3 <http://www.kb.se/kw3/ENG/Default.htm>

77 Aschenbrenner, A. Long-Term Preservation of Digital Material - Building an Archive to Preserve Digital Cultural Heritage from the Internet. <http://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando/Long_Term_Preservation.html> at <http://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando/Kulturarw3.html>

78 Arvidson, A.; Persson, K. & Mannerheim, J. The Kulturarw3 Project - The Royal Swedish Web Archiw3e - An example of "complete" collection of web pages. <http://www.ifla.org/IV/ifla66/papers/154-157e.htm>

79 Press Release, New decree for Kulturarw^3 <http://www.kb.se/Info/Pressmed/Arkiv/2002/020605_eng.htm>

3.2.3. The Nordic Web Archive (NWA)

The Nordic Web Archive (NWA)[80] is a forum in which all the Nordic National Libraries participate in order to co-ordinate and exchange information in the fields of harvesting and archiving web documents. Between November 2000 and July 2002 the NWA focused on developing software for accessing archived web documents. The outcome of this work was the NWA toolset,[81] a freely available solution for searching and navigating archived web document collections, built using PHP and Perl, and utilising open standards like the http protocol and XML for communication between different parts of the system. Use of the NWA toolset (i.e. searching and navigating a web archive) is done via a regular web browser, and no special plugins are needed to make it work. The NWA Toolset is to be released under the GNU General Public License early in 2003. The website contains no discussion of legal issues.

3.2.4. France - Bibliothèque de France

The National Library of France has carried out a series of studies relating to web archiving since 1998, building on the experience of other European States, with the aim of providing both a French web archive and sufficient information to allow for the updating of French legal deposit laws to take account of Internet publication.[82]

The National Library has considered the approaches taken by other states and divided them into two categories:

- manual selection and individual follow-up of the sites
- automatic collection

Neither of these approaches are seen as entirely satisfactory. The former permits the collection of web materials by direct deposit and overcomes some of the difficulties of on-line harvesting allowing for a high quality complete collection, but struggles to deal with non-traditional publishers, and often fails to collect the type of web materials most representative of the new medium. The latter permits the collection of a diverse range of web materials, and is relatively cheap in terms of staff time and costs, but cannot cope with restricted access web materials and databases (deep web as opposed to public web), and due to the sheer size of the collection, the number of 'snapshots' of the web may have to be limited.

The system proposed by the National Library of France combines elements of the two approaches, as well as elements drawn from the operation of web search engines such as Google. In broad terms, the system would use a search engine which would identify the location of French websites (those located on the national ccTLD '.fr', as well as those located in the gTLDS, com, net, org). the engine will then assign weight to each of the websites according to the number of links to them as well as other parameters used by current search engines. This will make it possible to automatically create a sample of French webpages on a nonarbitrary basis by selecting the most significant pages. The engine will also identify obstacles to collection encountered during the course of its search of the French Web (passworded sites, dynamic files, databases etc.). Using this information, the engine can provide details of deep web resources that are important, but not easily archived

---

80 The Nordic Web Archive <http://nwa.nb.no/>

81 About the NWA Toolset <http://nwa.nb.no/aboutNwaT.php>

82 Bibliothèque Nationale de France: Préparation de l'extension du dépôt légal de l'Internet à la Bibliothèque Nationale de France Approach BnF <http://www.bnf.fr/pages/infopro/depotleg/dli_intro.htm>

automatically, and this can be used to contact the owners of the materials, with a view to the materials being deposited after individual follow-up. This approach provides two key advantages:

The failure of automatic harvesting to cope with deep web resources is compensated for by targeted individual follow up for those resources which score highly on the 'number of links to' and other parameters.

The narrowness of coverage afforded by manual selection is overcome thanks to the engine which ensures that important websites are harvested or flagged for manual selection, but which also permits a broad sampling which can be filed automatically.

The National Library of France, like the other web archives examined, does not deal with issues outside copyright and legal deposit.

3.3. Future Developments

All three of the web archives examined above will be affected to some extent by their national implementations of the *European Copyright Directive.* At present it is not possible to state with any certainty what the effects might be, although it is possible that protections provided to digital rights management mechanisms in the Directive may make automatic harvesting a more hazardous task, particularly if, as Kulturarw[3] has suggested, future harvesting might ignore robots.txt and other site-based instructions/limitations on what may be acquired and indexed. Equally, changes to the *Data Protection Directive* or to national implementations or administrative practice may also affect automatic harvesting operations.

In general, however, in the three states examined above, there appears to be a clear recognition by governments that attempts to preserve national web-based materials are beneficial in nature, and in most cases, legal deposit rules have been altered, or are going to be altered, to facilitate those aims. From the materials available, it seems that none of the projects currently appear unduly concerned with legal issues beyond copyright and legal deposit and thus appear to have no strategies for dealing with issues such as defamation or obscene materials - this may seem short sighted, although it is quite possible that these issues have been considered and simply deemed to be of very low risk to what are, in essence, government approved (and often sponsored) projects.

4. The United States

It is perhaps unsurprising that the US, having led the way in terms of developing, and then commercialising, the Internet, would have one of the oldest and largest operations dedicated to archiving websites, in the form of the San Francisco-based Internet Archive. What is perhaps surprising is that, despite there being a relatively well developed body of law relating to the Internet in the US, relatively little study or analysis appears to have been undertaken/published with regard to the potential legal liabilities, beyond the basic issues of legal deposit of digital materials and copyright. It may be that, as far as existing US archives are concerned, explicit and public consideration of the wider legal issues pertaining to their activities is a can of worms that is most safely left unopened.

4.1. Legal Issues

Consideration of the legal issues in the US (as in Australia, below) is complicated by its federal nature, with both federal and state laws pertaining to some Internet activities. This is particularly true in terms of illegal content, where differing standards between the states make the area a potential minefield, with material that is acceptable in one state effectively criminalised in another.[83] That having been said, the extent of the First Amendment's protection of freedom of speech, as expanded by the jurisprudence of the Supreme Court, means that issues such as defamation and privacy have much narrower scopes than in the EU/UK.

4.1.1. Copyright

US copyright law is similar in many respects to that of the UK in that the authors of "original works of authorship," including literary, dramatic, musical, artistic, and certain other intellectual works are granted certain rights in those works.[84] This protection is available to both published and unpublished works in fixed form. No publication or registration or other action in the Copyright Office is required to secure copyright,[85] and the use of a copyright notice is no longer required under U. S. law. The owner of copyright thus has the exclusive right to do and to authorize others to do the following:

- reproduce the work in copies or phonorecords;
- prepare derivative works based upon the work;
- distribute copies or phonorecords of the work to the public by sale or other transfer of ownership, or by rental, lease, or lending;
- perform the work publicly, in the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works;
- display the copyrighted work publicly, in the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including the individual images of a motion picture or other audiovisual work.[86]

---

83 See, for example, *US v Thomas* F.3d 701 (6th Cir. 1996). See further below.

84 Title 17, U.S. Code

85 Although there are additional benefits to registration in the US. See further, US Copyright Office, 'Copyright Basics' <http://www.copyright.gov/circs/circ1.html>.

86 Copyright Act 1976, s.106.

Additionally, certain authors of works of visual art also have rights of attribution and integrity.[87] It is illegal for anyone to breach any of the rights provided by copyright law to the owner of copyright.

Copyright - Legal Deposit

All works under copyright protection and published in the United States are subject to mandatory deposit.[88] The mandatory deposit provision ensures that the Copyright Office is entitled to receive copies of every copyrightable work published in the United States. These deposits "are available to the Library of Congress for its collections, or for exchange or transfer to any other library."[89] The Act requires the "owner of copyright or of the exclusive right of publication" in a work published in the United States to deposit the required number of copies in the Copyright Office within 3 months of the date of such publication. . Over the years, the law has changed to take into account various new categories of material, for example, in 1989, previously exempt computer programs and “data” published in machine-readable copies (e.g. CD-ROMs) became subject to mandatory deposit by regulation as the best edition of a work. However, deposits of networked electronic publications solely available online are not presently required by regulation at this time.

> *The Copyright Office of LC takes a broad view of deposit categories and considers all types of publications subject to mandatory deposit but the Library of Congress’s current best edition requirements do not yet cover networked publications available only online. The questions of what constitutes “publication”, “transmission”, and “copies” when copyrighted works in digital form are made available only online present complex legal issues which must be resolved and applied in the context of mandatory deposit.*[90]

Copyright - Library and Archive Copying

Until 1998, US copyright law allowed libraries and archives to reproduce and distribute one copy of a work under certain circumstances[91] e.g. photocopies for interlibrary loan,[92] and to make copies for preservation purposes.[93] Following the Digital Millennium Copyright Act of 1998, a library may now make up to three copies (instead of one copy) of an unpublished work for purposes of preservation, including copies in digital form as long as that format is not made available to the public outside of the library or archives. A library may also make up to three copies (instead of one copy) of a published work to replace a damaged, deteriorating, lost, or stolen work (when an unused replacement cannot be obtained at a fair cost). The library may also make up to three digital copies to replace a work in an obsolete format as

---

87 Copyright Act 1976, s.106A.

88 US Copyright Office, Mandatory Deposit of Copies or Phonorecords for the Library of Congress <http://www.copyright.gov/circs/circ07d.html>

89 Copyright Act 1976, s.407.

90 Martin, E. ‘Management of Networked Electronic Publications - A Table of Status in Various Countries (revised 2001), National Library of Canada <http://www.nlc-bnc.ca/obj/r7/f6/r7-100-e.rtf>

91 Copyright Act 1976 s.108

92 Copyright Act 1976 s.108b

93 Copyright Act 1976 s.108c

long as that format is not made available to the public outside of the library or archives.[94] As with UK, however, the wording of the statute is such that it would clearly not cover the process of web archiving inasmuch as the library or archive does not already own a copy of the work.

Copyright - Licensing

As with the UK, in the absence of a statutory right to archive webpages, either by means of legal deposit, or under other library and archive privileges, it would appear that the only way that the webpages in the US can be archived in full conformity with copyright law is for the would-be archivist to endeavour to obtain the necessary permissions to copy the works in particular webpages from the relevant rightholders.

4.1.2. Defamation

Whilst there are great similarities between the common law of England and the US in respect of defamation there are also significant differences (e.g. under English law, in cases of innocent dissemination, the defendant publisher has to establish his innocence, whereas under American law the plaintiff who has been libelled has to prove that the publisher was not innocent). Thus, while we can say that the general thrust of defamation law is similar in the two countries, direct comparisons are often difficult.

Defamation law in the United States has changed substantially over the years, but is now well-established. For a plaintiff to prevail in a US defamation action, he must prove publication of the defamatory statement, identification of the plaintiff, falsity, defamatory content, injury and fault. If the plaintiff is a public official or public figure and the subject matter is a matter of public concern, or if the plaintiff is a private individual seeking punitive damages for a statement involving a matter of public concern, he must prove actual malice to establish the fault element.

Defamation law with relation to the Internet and particularly with regard to Internet Service Providers and Internet Intermediaries does throw up some signal differences, as the first general immunity provisions for ISPs were introduced by the US, via the Communications Decency Act (CDA) of 1996. The provisions which provide that immunity, were not in fact supposed to be the main thrust of the Act. It drafters intended the CDA to introduce new criminal offences of knowingly creating, sending, transmitting or displaying obscene or indecent materials to minors, or knowingly permitting the use of one's telecommunications systems for these purposes. The ISP immunity provisions in §230 were added to overrule an earlier decision[95] which had made it risky for an ISP to exercise any monitoring of the content it carried, such as introducing blocking or filtering technology, without rendering itself potentially liable as an editor or publisher, and thus becoming responsible for any third party content that it carried. The drafters intended ISPs who acted as "Good Samaritans" by 'protecting' their users from obscene or indecent materials using such technologies, to escape any resulting liability.[96]

---

94 US Copyright Office, Reproductions of Copyrighted Works by Educators and Librarians <http://www.copyright.gov/circs/circ21.pdf>

95 *Stratton Oakmont, Inc. v. Prodigy Services. Co.* 23 Media Law Rep. (BNA) 1794, 5 CCH Computer Cases ¶ 47,291 (N.Y.Sup.Ct. 1995).

96 The Senate conference report on § 230 states:
"This section provides "Good Samaritan" protections from civil liability for providers or users of an interactive computer service for actions to restrict or to enable restriction of access to objectionable online material. One of the specific purposes of this section is to overrule *Stratton-Oakmont v. Prodigy* and any other similar decisions which have treated such providers and users as publishers or speakers of

Regrettably for the CDA's drafters, the new criminal offences were struck down in ***A.C.L.U. v. Reno***[97] as the Court felt its "indecent transmission" and "patently offensive display" provisions abridged "the freedom of speech" protected by the First Amendment as it lacked the precision that the First Amendment requires when a statute regulates the content of speech. However, the striking down of the criminal provisions left the immunity provisions untouched. §230 has been tested a number of times in subsequent litigation, but at present the US courts appear to consider that it provides complete immunity from civil actions for defamation,[98] even where the ISP pays the author for the right to provide access to the defamatory material,[99] and it has also been successfully used as a defence in a civil action alleging negligence in failing to prevent continued solicitations to purchase child pornography made via an ISP's system.[100]

---

content that is not their own because they have restricted access to objectionable material. The conferees believe that such decisions create serious obstacles to the important federal policy of empowering parents to determine the content of communications their children receive through interactive computer services." S. Conf. Rep. No. 104-230, at 435 (1996).

97 929 F. Supp. 824, 830-838 (E.D. Pa. 1996), **affirmed** 117 S. Ct. 2329 (1997).

98 *Zeran v. America Online*, 129 F.3d 327 (4th Cir. 11/12/97), affirming 958 F. Supp. 1124 (E.D. Va., 3/21/97), cert. denied. (ISP not liable for allegedly defamatory postings by one of its subscribers. Plaintiff maintained that ISP was negligent in permitting anonymous postings by a subscriber accusing plaintiff of publishing materials "glorifying" the Oklahoma City bombing. The superior court affirmed that the claim was pre-empted by Section 230(c)(1) CDA, immunizing Internet service providers from "distributor liability." Imposing liability would create a disincentive for providers to review content for potentially objectionable material, and would thus frustrate one of the CDA's chief aims. Also if liability were incurred merely upon ISP being notified of allegedly improper material, it would place a burden of investigation and judgement far greater than that on traditional print publishers -- an impossible burden in the Internet context, and a clear invitation to third parties to foment lawsuits and leverage settlements by merely sending notice and demanding action.); *Aquino v. ElectriCiti* 26 Media L. Rep. 1032 (San Francisco Super. Ct. 1997) (Plaintiff sued claiming ISP had failed to take action against a subscriber who sent messages to a Usenet group allegedly defaming plaintiffs by associating them with a previous child-abuse investigation. The court dismissed the action under the Communications Decency Act's "safe harbor" provisions for Internet Service Providers, citing the rule in *Zeran v. AOL*.); *Kempf v. Time, Inc*., No. BC 184799 (L.A. Supr. Ct. 6/11/98): (plaintiff's admission that defendant ISP played no role in creating or developing allegedly libellous material, meant action against the ISP dismissed under the "safe harbor" provisions of the Communications Decency Act); *Ben Ezra, Weinstein & Co. v. America Online Inc.*, No. CIV 97-485 (D.N.Mex. 3/1/99) (BW&C sued for defamation, alleging that AOL published incorrect and defamatory information on the Internet concerning BW&C's publicly traded stock. Court held that AOL "clearly qualifies" for Internet service provider immunity under § 230 of the Communications Decency Act of 1996).

99 *Blumenthal v. Drudge*, 992 F. Supp. 44 (D.D.C. 1998) (ISP shielded from liability as a content provider for allegedly defamatory statements about plaintiff written by columnist under contract to ISP and published on ISP's service. Court held following *Zeran v. America Online* that the safe harbor provisions of the CDA absolutely precluded state common law defamation actions against Internet service providers).

100 *Doe v. America Online Inc.*, No. 25 Media L. Rep. 2112 (Fl. Cir. Ct. 1997). (ISP liability shield provision of CDA protected ISP from liability for subscriber's use of chat room to advertise pornographic images of 11-year-old boy).

Thus, in the US ISPs clearly have a general immunity from liability for defamatory statements made by others. The question is, will the courts be inclined to stretch this immunity to cover institutions such as web archives? It is arguable that a general harvesting website like the Internet Archive engages in little or no editorial selection of the content in its archive, but it would appear to be more than a 'passive conduit' for data, as ISPs are increasingly considered. This issue remains to be decided by the courts.

4.1.3. Data Protection

In the United States, the approach taken to the concept of personal data privacy is a somewhat complex one. Despite the lack of an explicit Constitutional basis for a right to privacy, the concept of privacy in the sense of "the right to be let alone"[101] has long been accepted in principle by the US legal system as a constitutional right, if rarely enthusiastically supported in practice with regard to informational privacy.[102] In fact, of the Bill of Rights, the First, Third, Fourth, Fifth, Ninth, and Fourteenth Amendments all contain elements attributable to a right to privacy.[103] Yet the types of privacy issues addressed by federal and state legislators and courts have tended to revolve around physical privacy[104] and decisional privacy.[105] Additionally, those constitutional privacy rights are always exercised against either federal, or state, government. Constitutional rights prevent the government from encroaching upon an individual's rights, they do not require the government to protect those rights against third parties.[106] Thus, records held by third parties are usually not protected unless there is specific legislation, and even then that legislation may be subject to challenge under the First Amendment.[107]

The US does not lack personal data privacy laws outside the constitutional sphere, as a scan through *The Privacy Law Sourcebook 1999*[108] amply demonstrates. Here one finds fourteen federal laws with some personal data privacy element - adding state laws and regulations would create a list running into the hundreds.[109] What the US lacks is both a coherent personal

---

101 The phrase drawn from the seminal article by Brandeis, L.D. & Warren S. "The Right to Privacy, the Implicit made Explicit" 4 (1890) *Harvard Law Review* 193.

102 For an excellent discussion of the historical and philosophical development of privacy theory in US law, see further Scoglio, S. *Transforming Privacy: A Transpersonal Philosophy of Rights* (New York: Praeger 1998). Also Schwartz, P.M. & Reidenberg, J.R. *Data Privacy Law* (Charlottesville: Michie 1996).

103 Scoglio, *Transforming Privacy*, *Ibid.* at 226 and *Griswold v. Connecticut* 381 U.S. 479 (1985).

104 See *Katz v. U.S.*, 386 U.S. 954 (1967).

105 See *Roe v. Wade*, 410 U.S. 113 (1973).

106 Cate, F. H., *Privacy in the Information Age* (Washington D.C.: Brookings Institution Press, 1997) at p. 99.

107 See further Carroll, M.W., "Garbage in: Emerging Media and Regulation of Unsolicited Commercial Solicitations" *Berkeley Technology Law Journal* Volume 11: Issue 2, Fall 1996. <http://www.256.com/~gray/spam/law.html>.

108 Rotenberg, M. *The Privacy Law Sourcebook 1999* (Washington DC: EPIC, 1999).

109 See Smith, R.E. (ed.) *Compilation of State and Federal Privacy Laws* (Privacy Journal, 1997) and EPIC's *Privacy Laws by State* at <http://www.epic.org/privacy/consumer/states.html>

data privacy framework, and any meaningful enforcement mechanism. As Rotenberg notes,[110] US federal privacy statutes have tended to arise less out of a concerted attempt to provide US citizens with a coherent personal data privacy regime, than out of a series of attempts to either fill legal lacuna that the courts had specifically refused to address,[111] or to assuage public concern arising from the use and abuse of new technologies.[112] The same is largely true of the efforts of the state legislatures.

The most heavily regulated sector in the US with regard to data privacy remains the government. Not only are there important constitutional controls on its ability to collect and use personal data in the law enforcement sector, but with regard to government collection and use of personal data for other purposes, most aspects of federal agency collection, maintenance, use and disclosure of personal information is regulated by the Privacy Act 1974.[113]

The US legal system thus recognizes a fundamental right of personal privacy, but it is clear that this right becomes rather nebulous in the area of informational privacy. The federal legislation fails to provide a comprehensive regime for data privacy, and the state coverage is similarly patchy. Even where federal legislation exists it is often so laden with exemptions as to virtually negate its purpose. This is due largely to the fact that much of the legislation appears to have been designed primarily to send a message from the legislature, either of reassurance to the voters, or of rebuke to the courts, without the measure actually impinging unduly on the ability of either government or commerce to maintain the *status quo.*

As such, a US web archivist is unlikely to be unduly troubled by data privacy considerations with regard to the harvesting of webpages containing personal data to the same degree as his/her European counterpart.

4.1.4. Illegal Content

In the US, 'obscenity' is limited to sexual material, and requires the material to appeal to the prurient interest, as defined by reference to the standards of the local community, and to depict sexual conduct defined by the applicable State law. The three-part test set out by the Supreme Court is:

> *(a) whether the average person, applying contemporary community standards, would find that the work, taken as a whole, appeals to the prurient interest,*
>
> *(b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law, and*
>
> *(c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value.*[114]

This test is not based on the potential effects of the material, but on whether it contravenes locally determined standards of acceptable sexual depiction. This leads to the somewhat

---

110 Rotenberg, *The Privacy Law Sourcebook 1999,* op.cit. n.108 at i-ii.

111 E.g. Right to Financial Privacy Act 1978; Privacy Protection Act 1980.

112 E.g. Privacy Act of 1974, 5 USC 552a; Video Privacy Protection Act of 1988, Telephone Consumer Protection Act 1991.

113 5 U.S.C. 552(a) <http://www.usdoj.gov/foia/privstat.htm>.

114 *Miller v. California* 413 US 15 (1973).

unfortunate result that material which is unobjectionable in one US state may be viewed as obscene in another, with potentially deleterious effects for the publishers. In the traditional media, publishers can largely avoid falling foul of locally determined standards, by adjusting their distribution networks accordingly. For a web archive, this distribution control approach may be untenable, as those using or accessing a potentially objectionable archived webpage might be based anywhere in the US.[115] However, some degree of age-based access control can be exerted through a variety of means, including the use of credit card validation, and an age-based access control system combined with disclaimers including a statement that the archive is merely a repository for information and the archivists do not exert editorial control over the contents might be sufficient to protect a general harvesting web archive. Where archivists do select the material to be archived, and potentially obscene material may be selected, further serious thought will need to be given to the nature of the access controls.

There have been several recent attempts to pass US federal legislation relating to Internet content control which would have imposed criminal liability, or other sanctions, on the provision of access to pornographic material on the Internet, including the Communications Decency Act 1996, the Child Online Protection Act 1998, and the Children's Internet Protection Act 2000. Thus far, none of these have withstood judicial scrutiny; in the main due to a failure to convince the judiciary that the ways in which the laws have been drafted do not unduly interfere with First Amendment rights of free speech.

4.2. Existing Archives and Policies

There appear to have been two major Internet archiving projects in the US. One, the MINERVA Project, was a pilot scheme run by the US Library of Congress to assess how it might most effectively collect and preserve materials from the Web. The other, the Internet Archive, a not-for-profit organization based in San Francisco has been collecting all open access HTML pages, approximately monthly, since 1996. A commercial company, Alexa Internet, carries out the data gathering and donates the data collected to the Internet Archive when it is six months old.

4.2.1. Library of Congress - Minerva

The MINERVA[116] Web Preservation Project was established to initiate a broad program to collect and preserve open access Web materials that the creators have made publicly available, without restriction, and which can be collected by simply downloading them over the Internet. The purpose of the MINERVA prototype was to gain insights into the practical issues involved in collecting and organizing selected Web sites, and to understand how the Library of Congress might operate a full-scale preservation program. The MINERVA project followed a selective collection strategy rather than a bulk collection strategy.

The main activities of the MINERVA project were:

- The selection of a small number of Web sites for close study.

---

115 This problem is clearly demonstrated by the case of *United States v. Thomas* 74 F.3d 701 (6th Cir.), Cert. denied, 117 S. Ct. 74 (1996), where a bulletin board operator was extradited from California to Tennessee to face criminal charges. It was stated in argument that the material, which was stored on a computer in California, was not obscene by Californian community standards, but the court determined that the appropriate standards by which to test for obscenity were the standards of Tennessee, the place in which the material was received and viewed.

116 Mapping the INternet: the Electronic Resources Virtual Archive

- The downloading of snapshots of the nominated sites which were inspected for errors, anomalies, etc.
- The creation of catalog records for the material collected
- The development of a trial Web site to demonstrate user access
- Discussion with the U.S. Copyright Office on legal issues.[117]

With regard to the legal situation for downloading open access materials the project felt that it was reasonable to assume that most organizations making information openly available on the Web would be willing for the Library of Congress to download copies and keep them for future research, and that the collection of born-digital materials for the benefit of future scholarship was clearly consistent with the Library of Congress's existing powers under the US Copyright Act[118] but noted that the Library of Congress did not at that point have the explicit legal right to do so.

To develop a full program of collecting and preserving open access Web sites, in this manner, it was felt that the Library of Congress needed legal authority for three additional activities:

- where materials have been made openly available without restrictions, to download copies from the Web rather than demand copies from the publisher, and to do so without having to ask permission before downloading.
- to designate one or more other organizations at separate locations to act as its agents to carry out collection and preservation of open access materials on its behalf.
- to make small editorial changes to the materials that it downloads for reasons of access and preservation.

The MINERVA project does not appear to have dealt with legal issues outside those of copyright and legal deposit in any detail.

4.2.2. The Internet Archive

The Internet Archive[119] is a 501(c)(3) public nonprofit organization[120] whose benefactors include Alexa Internet,[121] AT&T Research, Compaq, the Kahle/Austin Foundation, Prelinger

---

117 Arms, W.Y.; Adkins, R.; Ammen, C. & Hayes, A. 'Collecting and Preserving the Web: The Minerva Prototype', *RLG DigiNews* 5 (2) April 15 2001.

118 See s.407 US Code <http://www.law.cornell.edu/copyright/copyright.act.chapt4.html#17usc407>

The Library of Congress receives a copy of essentially all materials registered for copyright and has the right to demand copies of all materials published in the USA to add to its collections.

119 The Internet Archive <http://www.archive.org/about/about.php>

120 IRS Section 501(c)(3) is the section of the US tax code that defines nonprofit, charitable, tax-exempt organizations; 501(c)(3) organizations are further defined as public charities, private operating foundations, and private non-operating foundations. This status may explain why the Internet Archive seems to take fairly low key approach to influencing debate in the area of legal issues of archiving, as tax-exempt 501(c)(3) nonprofits are prohibited from acting to influence legislation

Archives, Quantum DLT, Xerox PARC, the Library of Congress, and the National Science Foundation. It was set up to provide permanent access for researchers, historians, and scholars to historical collections that exist in digital format. It was founded in 1996 and receives data donations from Alexa Internet and others.

As the archive is designed to preserve all publicly accessible materials displayed on the Internet, it operates an automated harvesting model, receiving its material from Alexa Internet after a delay of 6 months. As one commentator has noted:

> *Since their goal is to archive the public space of the internet, there is no real selection criterion. The [collection] policy is more geared towards what is* not *included, than what* is *included*.[122]

This approach has important implications for the legal situation as it pertains to the archive. The Internet Archive does not preserve the 'deep web' and webpages that are password protected, on private servers, or whose owners request not to be crawled are thus not archived. Additionally, the Internet Archive provides webpage owners with the ability to ask the Archive to remove any information collected from their site, and information on how webmasters can write simple html code to prevent robots from crawling their pages.[123]

The difficulty for Internet Archive is that regardless of the *a priori* or *a posteori* opt-out mechanisms it provides for webpage owners, as noted above in the discussion of UK copyright law, those mechanisms are in essence irrelevant to a copyright holder intent upon enforcing his copyright law rights via the courts. The Internet Archive has no legal deposit role, or other statutory rights, to permit it to archive the copyrighted material of others without permission, and is essentially dependant upon the goodwill of rightholders. This leaves the Internet Archive in a very weak position - it may be that the 6 month delay before material becomes available to the archive, in combination with the opt-out mechanisms, prevents serious legal challenges - however, this places a serious burden on the Internet Archive in terms of determining whether, and whose, rights have been infringed. A classic example of this has been the relatively recent incident concerning the Church of Scientology and its critics' websites. Here, the Church of Scientology's lawyers approached the Internet Archive suggesting that material on one of its critics' websites infringed the Church's intellectual property (copyright and trademark). The Internet Archive responded by removing the websites from the Archive and posting a notice stating:

> *Per the request of the site owner, http://www.xenu.net/ is no longer available in the Wayback Machine. Try another request or click here to see if the page is available, live, on the Web.*

There were a couple of problems with this approach. Firstly, the owner of the website blocked had not made any such request for material to be removed; secondly many of the webpages on

---

"except to an insubstantial degree". Brewster Kahle has, however, played a role in the debate and legal action surrounding the recent copyright term extension in the US.

[121] Alexa Internet <http://www.alexa.com/>

<http://pages.alexa.com/company/index.html?p=Dest_W_t_40_B1>

In June 1999, Alexa Internet became a wholly owned subsidiary of Amazon.com.

[122] Zimmerman, A. Digital Preservation Case Study: The Internet Archive <http://www.geocities.com/azitiz/paper_internetarchive.htm>

[123] Removing Documents From the Wayback Machine <http://www.archive.org/about/exclude.php>

the site blocked appear to have contained no infringing material.[124] At this point, however, it appears that although the posted notice has changed, the website remains blocked.

> *http://www.xenu.net is not available in the Wayback Machine. Try another request or click here to see if the page is available, live, on the Web. http://www.xenu.net*

This example, part of a much wider and often vitriolic battle between the Church of Scientology and its critics,[125] neatly demonstrates the Internet Archive's dilemma - because its harvesting may very well be a massive copyright violation, it is obliged to take seriously any allegation of infringement of copyright, regardless of the veracity of the claims made to it, and the most obvious approach to this difficulty is simply to remove any potentially offending material. This does not, as critics have pointed out, bode well for the Archive's stated aim of being "an 'Internet library,' with the purpose of offering permanent access for researchers, historians, and scholars", if controversial material can be so easily stifled by mere allegations of illegality.

As regards issues of access to the material collected, in theory the Internet Archive makes the collection available to the general public upon application[126] requiring an end user to agree to the Internet Archive's Terms of Use.[127] In practice, it seems to be possible to access large parts of the web archive as an anonymous user without encountering any formal means of access control or reading the Terms of Use.

The Terms of Use Agreement advises users that they use any content contained in the archive at their own risk, and requires them to agree to

> *…abide by all applicable laws and regulations, including intellectual property laws, in connection with your use of the Archive. In particular, you certify that your use of any part of the Archive's Collections will be non-commercial and will be limited to noninfringing or fair use under copyright law. In using the Archive's site, Collections, and/or services, you further agree*
>
> *(a) not to violate anyone's rights of privacy,*
>
> *(b) not to act in any way that might give rise to civil or criminal liability,*
>
> *(c) not to use or attempt to use another person's password,*
>
> *(d) not to collect or store personal data about anyone,*

---

124 Bowman, L.M. 'Net archive silences Scientology critic', news.com (24 Sept, 2002) <http://news.com.com/2100-1023-959236.html>

Wood, M. 'Church, DMCA, and too many missing links' cnet.com (27 Sept 2002) <http://www.cnet.com/software/0-8888-8-20472178-1.html>

125 See, for example, Gallagher, D.F., 'Google Runs Into Copyright Dispute' New York Times, (April 22, 2002) <http://www.nytimes.com/2002/04/22/technology/ebusiness/22NECO.html> (subscription required)

126 Get a Virtual Library card <http://www.archive.org/account/login.createaccount.php>

127 Internet Archive Terms of Use (10 March 2001) <http://www.archive.org/about/terms.php>

> *(e) not to infringe any copyright, trademark, patent, or other proprietary rights of any person,*
>
> *(f) not to transmit or facilitate the transmission of unsolicited email ("spam"),*
>
> *(g) not to harass, threaten, or otherwise annoy anyone, and*
>
> *(h) not to act in any way that might be harmful to minors, including, without limitation, transmitting or facilitating the transmission of child pornography, which is prohibited by federal law and may be reported to the authorities should it be discovered by the Archive.*

It is difficult to see how the archive might actually enforce these terms and conditions in practice, particularly as large parts of the Archive (including pornographic material) appear to have no formal access criteria or other means of identifying end-users. As with the Internet Archive's approach to the legalities of its collection of material, its approach to legal liability in this regard appears to be largely one of presentation, rather than one of any real substance.

The Terms of Use also warns users that:

> *Because the content of the Collections comes from around the world and from many different sectors, the Collections may contain information that might be deemed offensive, disturbing, pornographic, racist, sexist, bizarre, misleading, fraudulent, or otherwise objectionable. The Archive does not endorse or sponsor any content in the Collections, nor does it guarantee or warrant that the content available in the Collections is accurate, complete, noninfringing, or legally accessible in your jurisdiction, and you agree that you are solely responsible for abiding by all laws and regulations that may be applicable to the viewing of the content.*

In short, the Internet Archive largely ignores copyright law in the process of collecting its material, provides only a limited (and, arguably, effectively valueless) protection for the material once stored, and in effect disclaims any responsibility for what is done with the material by the end user, as well as any liability that the end user may incur in accessing the material. Given the litigious nature of the US, it will be interesting to see if the Internet Archive's success in avoiding litigation over its activities will continue for much longer.

4.3. Future Developments

It seems likely from the work undertaken by the MINERVA project that the Library of Congress will push for an expansion of the legal deposit system in the US to include digital materials such as webpages, and that such an expansion is likely to be granted, although whether there will be penalties for failing to comply with legal deposit obligations for webpages as there are for print works, must be in doubt. US copyright law is not, at present, particularly helpful for the non Library of Congress web archivist, and there must be some doubt, given the aggressively pro-content provider stance taken in recent copyright legislation such as the Digital Millennium Copyright Act 1998, whether this is likely to change in the near future.

While the issue of whether defamation law will be applied to Internet archives in the same way that it is to ISPs remains unclear (although the rationales for doing so are reasonably compelling), it seems that neither defamation law nor data privacy laws are going to be as significant a burden for US Internet archivists as they are for those archivists in the UK. Illegal content may pose problems, although the already explicit nature of many legal US websites, combined with First Amendment protections, suggests that US based Internet archives are likely to be largely shielded from legal difficulties.

5. Australia

As will become clear below, Australia has made great strides in the area of Internet archiving, led by the National Library of Australia. This is despite the fact that the Australian legislatures at both Federal and State levels appear to have been loathe to act in support of archiving initiatives. Very few of the laws potentially affecting web archiving have been updated in recent years. As regards the topic of legal deposit, while several of the Australian States have legislation that requires electronic publications to be deposited, the Australian Commonwealth *Copyright Act 1968* has yet to be updated to take into account the deposit of digital works. The National Library has been active in its efforts to achieve reform of legal deposit provisions.

5.1. Legal Issues

In very general terms, Australian law is roughly comparable to the law of the UK in the areas discussed below. The main difference lies in the Australian federal system of government, which means that as well as national federal laws, the would-be archivist must also consider the law of his or her Territory or State.

5.1.1. Copyright

The law of copyright in Australia is governed by the Copyright Act 1968 and the subsequent decisions of courts. As in the UK, protection of a work is free and automatic upon its creation. The term of copyright in Australia is shorter than in the UK and US at the author's life + 50 years. As Australia is a signatory to the Berne Convention, most foreign copyright owners are protected in Australia, and, while containing some differences, Australia's copyright law regime largely resembles that of the UK and US.

Australian law contains provisions imposing criminal penalties and civil remedies for making importing or commercially dealing in devices and services which circumvent technological copyright protection measures, and sanctions against tampering with electronic rights management information and against distributing or commercially dealing with material whose rights management information has been tampered with.[128]

5.1.2. Defamation

The law of defamation in Australia is complicated by the fact that Australian defamation laws are primarily State and Territory laws, rather than Federal laws, and that the law, including available defences, is different in each jurisdiction. Australian laws include offence provisions for civil defamation and criminal defamation. Civil liability arises from publications likely to harm a person's reputation and penalties are monetary. Criminal liability arises from publications that affect the community, such as those that have a tendency to endanger the public peace, and penalties in most jurisdictions include imprisonment. There are significant differences between civil and criminal defamation law relative to liability and defences.

The definition of "defamatory matter" varies among Australian jurisdictions. In some jurisdictions common law definitions apply, while in others, such as Queensland and Tasmania, the definition has been codified. In broad terms the definition is similar to that of defamation law in the UK, in that statements that would lower the reputation of an individual in the eyes of others are potentially defamatory. Similarly, as in the UK, if a defamation action is to be successful, it must be established that the communication:

---

[128] See further Australian Copyright Council Online Information Centre <http://www.copyright.org.au/>

- was published to a third person, i.e. to at least one person other than the plaintiff
- identifies the plaintiff, for example, by name or by a reference to a small group of people.
- contains a defamatory statement or imputation (whether intentionally published or not).

Defences that may be successfully pleaded in relation to a defamation action vary throughout Australian jurisdictions. Depending on the jurisdiction, these may include:

- truth/justification
- fair comment
- absolute privilege (this attaches to the occasion, not the statement or speaker, such as during parliamentary proceedings, judicial and quasi-judicial proceedings, executive communications and communications between spouses)
- qualified privilege (e.g. fair and accurate reports of parliamentary proceedings, judicial proceedings, public meetings concerning matters of public interest/concern)
- consent (e.g. where the plaintiff expressly or impliedly consented to the publication of the particular imputation)
- triviality (e.g. where the circumstances/occasion of the publication were trivial to the extent that the person defamed was not likely to suffer harm)
- innocent dissemination (e.g. applicable to re-publishers/re-distributors such as newsagents/book sellers, including potentially to ISPs/ICHs.

The circumstances in which the above defences may be applicable varies among Australian jurisdictions. For example, truth alone is not a defence in all jurisdictions. In some, the defendant must also prove that the publication of a true statement or imputation was made for the 'public benefit' or relates to a matter of 'public interest'.

The Federal Broadcasting Services Act 1992[129] provides a statutory defence to an Internet Service Provider or Internet Content Host who carries/hosts Internet content in Australia and who was not aware that they were carrying/hosting a defamatory publication. s.91(1) of Schedule 5 to the Broadcasting Services Act provides that a law of a State or Territory, or a rule of common law or equity, has no effect to the extent to which it:

> *(i) subjects, or would have the effect (whether direct or indirect) of subjecting, an internet content host/internet service provider to liability (whether criminal or civil) in respect of hosting/carrying particular internet content in a case where the host/provider was not aware of the nature of the internet content; or*
>
> *(ii) requires, or would have the effect (whether direct or indirect) of requiring, an internet content host/internet service provider to monitor, make inquiries about, or keep records of, internet content hosted/carried by the host/provider.*

The definition of "internet content" in the BSA excludes "ordinary electronic mail", information that is transmitted in the form of a broadcasting service and information that is not "kept on a data storage device". Hence, the s.91 defence will not be available in cases

---

[129] <http://scaleplus.law.gov.au/html/pasteact/0/136/top.htm>

involving such material. In these cases, an Internet Service Provider or Internet Content Host may be able to rely on the defence of innocent dissemination. The common law defence of innocent dissemination has historically applied to re-distributors such as newsagents, booksellers, libraries, etc. An ISP or ICH may also be able to rely on the common law defence of innocent dissemination in circumstances where they did not know that the publication was defamatory or likely to contain defamatory matter and their absence of knowledge was not due to negligence on their part. Whether the common law defence of innocent dissemination can be relied upon by ISPs and ICHs has not however yet been determined by the Australian courts.[130]

5.1.3. Data Protection

Australia has a relatively new informational privacy regime at the federal level based on the *Privacy Act 1988* which initially applied mainly to Commonwealth and ACT Government public sector agencies. In December 2000, the *Privacy Amendment (Private Sector) Act 2000* amended the Privacy Act (with effect from 21 December 2001) and it now applies to many private sector organisations as well. In many ways the Australian legislation resembles that of the EU Member States, both being based originally upon the Organisation for Economic Cooperation and Development's (OECD) Guidelines on the Protection of Privacy and Transborder Flows of Personal Data developed in 1980, and both applying to both public and private sectors. The National Privacy Principles (the NPPs) in the Privacy Act set out how private sector organisations should collect, use, keep secure and disclose personal information. The principles give individuals a right to know what information an organisation holds about them and a right to correct that information if it is wrong. It would seem that an Internet archive set up in Australia, whether a public or private body, would be covered by the amended Privacy Act. Additionally, there is a body of State and Territory privacy legislation that has to be considered.[131]

5.1.4. Content Liability

The primary Australian legislation in this area is the *Broadcasting Services Amendment* (*Online Services) Act* 1999 (Cth).[132] This provides three basic strategies:

- Censorship and content regulation
- The creation of industry codes of practice
- Community education

The law forbids the posting of prohibited material[133] and potential prohibited material[134] and requires the Australian Broadcasting Authority (ABA) to investigate websites after they have

---

130 Much of the above discussion is drawn from EFA, Defamation Laws & the Internet <http://www.efa.org.au/Issues/Censor/defamation.html>

131 See further, The Office of the Federal Privacy Commissioner, Privacy in Australia <http://www.privacy.gov.au/publications/pia1.html>

132 Incorporated as Schedule 5 of the *Broadcasting Services Act* (1992) (Cth).

133 Prohibited material is material that the Australian Office of Film and Literature Classification (AOFLC) would refuse classification (RC), rate X, or rate R, hosted in Australia and not within a restricted access scheme. RC rated material can be legally hosted in Australia, but only where the site has installed mechanisms to ensure adults-only access.

received complaints from consumers. The ABA does not actively monitor internet content for censorship purposes, and the law applies only to Internet companies based in Australia.

Where material complained about is prohibited or potentially prohibited, the ABA takes different action depending on where the material is housed. If it is hosted in Australia, the content host will be given a direction to remove the material from the site.[135]

For material hosted outside Australia, part of the legislation appears to suggest that it was intended that steps be taken to prevent internet users from accessing prohibited or potentially prohibited content, although it did not specify how this was to be done, merely stating that all reasonable steps should be taken to that end.[136] However, under the legislation, the fact that there is an Industry Code of Practice, written by the Australian Internet Industry Association (IIA) and registered by the ABA, means that ISPs are only obliged to comply with the industry code.[137]

The Industry Code of Practice states that ISPs will be taken to have complied with the legislation, in regard to overseas hosted content, where they provide a content filter, or filtered ISP service, for their users. Thus, ISPs notified of prohibited and potential prohibited material need not take steps to block it. There is no obligation on users to use the filters or filtered services.

Thus, while prohibited and potentially prohibited material may not be legally hosted in Australia, there is no regulation of any material hosted outside Australia.

5.2. Existing Archives and Policies

There is one overarching Internet archiving project in Australia known as PANDORA (Preserving and Accessing Networked DOcumentary Resources of Australia)[138] which is co-ordinated by the National Library of Australia. Project partners include The State Library of Victoria, State Library of South Australia, State Library of Western Australia, ScreenSound Australia, State Library of New South Wales, State Library of Queensland and the Northern Territory Library and Information Service all of whom are full partners in PANDORA, and are selecting, cataloguing and archiving publications in their areas of interest.[139]

5.2.1. National Library of Australia - PANDORA

In 1996, the National Library of Australia began building an archive of selected, significant Australian web sites and web-based online publications known as the PANDORA archive. Unlike automatically harvesting archives such as Kulturarw[3] and the Internet Archive, the content of PANDORA is produced via a highly selective process. This inevitably means that

---

134 Potential prohibited material is material likely to be refused classification (RC), rated X, or rated R.

135 Anon., Australian Internet Anti-Pornography Effort Accelerates But May Be Ineffective, March 27, 2000 <http://www.adlawbyrequest.com/international/AussieAnti-Porn.shtml>

136 *Broadcasting Services Act* (1992) (Cth) schedule 5 s.40(1) (c)

137 *Broadcasting Services Act* (1992) (Cth) schedule 5 s.40(1)(b)

138 The PANDORA Archive <http://pandora.nla.gov.au/index.html>

139 Background Information About PANDORA: the National Collection of Australian Online Publications <http://pandora.nla.gov.au/background.html>

the number of sites archived is limited, e.g. in June 2001, it contained just 1250 web sites. However, the aim of the archive is to obtain a strongly representative sample of Australian web publishing by academic, government, commercial and community organisations and to preserve material relating to historic events. Already a number of the web sites archived, including the official web site for the Sydney Olympic Games, have disappeared from the live Internet. About one-third of the 1250 web sites have been captured on multiple occasions, allowing the archive to build up a sequence of snapshots which demonstrate how these sites have changed over time.[140]

PANDORA is characterised by the following features:

- it is based on clear selection guidelines;
- web sites are gathered and managed using software called PANDAS (PANdora Digital Archiving System);
- the gathering is undertaken by partner institutions in addition to the Library itself;
- permission of the publisher is sought and received prior to any site being included in the archive;
- every web site is catalogued, and the catalogue entries are included in the Library's Catalogue, the National Bibliographic Database and PANDORA web site
- the archived version of every web site is subjected to quality checking to ensure that all files have been correctly captured.[141]

Access controls can be implemented within the PANDORA archive and access restrictions are applied to some archived sites, these restrictions are managed by the PANDAS system. Restrictions are applied for commercial reasons, for privacy or cultural reasons, or as part of a policy decision for certain categories of material.

- Commercially produced and distributed material selected for PANDORA is subject to a negotiation process between the Library and the publisher to determine access conditions that will not undermine the publisher's commercial viability.
- Some material of a sensitive nature may also be restricted. For such material suitable time restrictions are negotiated, or in some cases only password controlled access to the title by designated users is permitted.
- Sometimes no public access to a title may be allowed. In some cases where permission for access in PANDORA has been denied, a title is considered so significant that it has been captured in the hope that it can eventually be made available. Libellous or other legally questionable material may be restricted to staff access only.
- Some adult material has been selected for inclusion in PANDORA to reflect the widespread availability of such material on the Internet. While this material is publicly accessible on the Internet, access is password controlled to the archived versions in PANDORA.

---

140 Cathro, W.; Webb, C. & Whiting, J. Archiving the Web: The PANDORA Archive at the National Library of Australia <http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html>

141 *Ibid*.

PANDAS manages all of these restrictions and allows only those users in designated locations (based on IP address) or with the required password to access the archived versions. As time periods for restrictions expire, the system automatically updates the title entry pages to indicate the changed access conditions.[142]

It will be clear from the forgoing that PANDORA is one of the more advanced and sophisticated of the national Internet archives. It is also clear that significant thought has gone into structuring the collection of and access to, the archival material with the aim of avoiding incurring significant legal risk. The model adopted would seem eminently suitable for Internet archives wishing to preserve a limited number of subject specific websites, whether these are sites of national importance or of particular subject relevance. This is not to say that the PANDORA archive has succeeded in removing the threat of litigation altogether - the restriction to staff access of defamatory material may still leave the project open to an action for defamation, even if the material is not accessible to the general public.

5.3. Future Developments

In Australia, as in other Berne Convention countries, it seems likely that digital copyright protection will continue to expand at the expense of public oriented doctrines such as fair use. However, as PANDORA obtains permissions for its archived material, this does not appear to be a problem for the project. It remains to be seen whether the new privacy laws will have any significant effect on web archiving, current developments suggest they will not. The key legal threat to the project would appear to come from a *Loutchansky* type defamation case, although the project appears to be organised in such a fashion that rapid take-down of offending materials would not be a problem. There is little sign at present of any legislative effort, at either federal or state level, to harmonise defamation law in Australia.

142 *Ibid*.

6. Conclusion - Running an Internet Archive in the UK

With reference to the foregoing, it would seem that some lessons can be learnt from other jurisdictions, notably that in order to make any headway in terms of dealing with the current legal difficulties surrounding the archiving of webpages one needs to either:

- attract government support for the task of creating a large scale archive (as appears to be the case in France and the Scandinavian countries) perhaps by playing to nationalist sensibilities ('the history of the vibrant culture of our nation's on-line presence will be lost forever') and obtain legislative permissions for copying and legislative protections from defamation, content liability and data protection laws;

- begin the creation of a large scale web archive without legislative permissions and protections and gamble that a combination of operational precautions and favourable media coverage will be sufficient to deter or deflect the likelihood of individuals or regulators bringing legal actions that halt or disrupt the activities of the archive (the US *Internet Archive* is a classic example of this approach);

- engage in a less expansive web archiving projective involving selective archiving in negotiation with the "publishers" of the material (the Australian PANDORA project appears to be adopting this model). This would also have the advantage of potentially opening the way for inclusion of some of the 'deep Web' i.e. subscription or otherwise restricted material, in the archive.

All these options are open to a UK archive, depending upon the level of risk that the archive operator is willing to accept.

6.1. Risks

Waiting for UK government intervention in this sphere is likely to be a drawn-out and frustrating affair, as the legislative timetable does not contain much space for specialist topics that may be of little immediate interest to government ministers. In the interim, some valuable web-based information, be it historical, medical or cultural, is likely to be lost permanently.

If a UK-based web archive were to be set up, it could attempt the kind of approach taken by the US *Internet Archive*. This would involve:

- Careful publicity designed to show the value of the web archive for historical research etc. with the implicit message that to wilfully obstruct such a worthy project on base legal grounds would be the act of a money grubbing philistine or a techno-luddite.

- Provision of lots of reassurances to IP rightholders that their IP in works will be respected, including provision of a mechanism for opting out of collection of a website or webpage (e.g. by use of some agreed system of opt out such as Robots.txt), and a mechanism to allow rightholders to withdraw their works after they have been collected by way of notice to the web archivist.

- Provision of a mechanism by which individuals who feel that the archive contains material which defames them can notify the web archivist. The archivist can then decide whether a) to post a notice on the material stating that the truth of the material is contested, or that it is the subject of defamation proceedings b) to remove the information from the archive.

- Provision of a mechanism by which individuals (and law enforcement agencies) who feel that the archive contains material which is obscene or indecent can notify the web archivist. The archivist can then decide whether to a) remove the material, b) restrict access to it, or c) make the case that its open display is "in the public good".

- Provision of a mechanism by which individuals who feel that the archive contains material which breaches their data protection rights can notify the web archivist. The archivist can then decide whether to a) attach a statement about the accuracy of the data to the webpage, b) amend the page, or c) remove the page.

- A time lag between the collection of the material and its appearance in the archive. This allows a time period for rightholders to assert their rights and for legal difficulties with regard to content to become known. It also reduces the likelihood that most rightholders will suffer significant financial loss as a result of the display of copyright material in the archive.

It should be noted that none of these mechanisms completely removes the possibility of legal action against the web archive. With regard to copyright and data protection in particular the mechanisms suggested offer little, if any, protection from liability should an aggrieved individual choose take the matter to court.

6.2. Opportunities

On the evidence thus far, under the present legal regime in the UK, organisations such as JISC and the Wellcome Trust would perhaps be best served by a web archiving strategy like that adopted by the National Library of Australia's PANDORA project. Such a project, by negotiating with web publishers and rightholders prior to archiving, could significantly reduce its exposure to legal risk. The extent to which the risk would be reduced would depend on the extent to which the burden of establishing that the material to be archived would not breach copyright or defamation, content liability and data protection laws might be shifted to those providing the material. As noted above, an additional advantage to this approach is that it provides a model which would also allow for the archiving of 'deep Web' resources in agreement with publishers and rightholders, thus adding further value to the archive.

7. Recommendations

1. Even in the UK, where clearly a quite different legal regime would apply, the automatic harvesting approach taken by the US Internet Archive is perhaps not as risky as it might at first appear, if adequate administrative precautions are taken. However, JISC/Wellcome may well be seen as potential 'deep pocket defendants' by some parties. It would be possible to set up a limited company, either in the UK or abroad, to do the harvesting, which could be funded by JISC/Wellcome by donation, but this might be seen to be encouraging or abetting copyright infringement, something that neither JISC or Wellcome would wish to be associated with. **It is thus RECOMMENDED that, given the current legal situation in the UK that JISC/Wellcome do not adopt the US Internet Archive automatic harvesting approach.**

2. While the National Library of France model, combining automatic harvesting with targeted individual follow up, is attractive, under the current UK legal regime this approach too is potentially risky without some form of legal deposit scheme combined with legal protection for the archivist. **It is thus RECOMMENDED that, given the current legal situation in the UK that JISC/Wellcome do not adopt the National Library of France model.**

3. The most suitable model at present appears to be that of the National Library of Australia's PANDORA project which is selective archiving of websites premised on obtaining the permission of the relevant rightsholders in advance of the archiving process. This signally reduces the risk of copyright infringement to the archivist, and allows for content liability risk to be distributed between website owner and the archive according to contract. **It is thus RECOMMENDED that, given the current legal situation in the UK that JISC/Wellcome pursue a web archiving strategy based on the National Library of Australia's PANDORA project.**

4. It is clear that UK law relating to legal deposit of copyright materials is in urgent need of updating with regard to the preservation of digital materials. Without such an update, significant amounts of important digital material, including public and deep web resources will be lost to posterity because it is difficult or impossible to legally archive them under the existing UK legal regime. There are a number of avenues open to JISC/Wellcome to influence the debate in this area - the most obvious is to encourage legislative action, either by means of lobbying the government directly or alternatively by making representations to MPs to secure a Private Members Bill on the topic. **It is thus RECOMMENDED that JISC/Wellcome consider future strategy to obtain necessary changes in UK law to allow the legal deposit and /or archiving of UK digital materials, and particularly UK web materials, and that such strategy should include approaches to government and MPs noting the loss to the UK of potentially valuable social and historical materials due to the current gap in UK legislation.**

Appendix A - UK Legislation

Copyright, Designs and Patents Act 1988

[…]

42.—(1) The librarian or archivist of a prescribed library or archive may, if the prescribed conditions are complied with, make a copy from any item in the permanent collection of the library or archive—

> (a) in order to preserve or replace that item by placing the copy in its permanent collection in addition to or in place of it, or
>
> (b) in order to replace in the permanent collection of another prescribed library or archive an item which has been lost, destroyed or damaged,

without infringing the copyright in any literary, dramatic or musical work, in any illustrations accompanying such a work or, in the case of a published edition, in the typographical arrangement.

(2) The prescribed conditions shall include provision for restricting the making of copies to cases where it is not reasonably practicable to purchase a copy of the item in question to fulfil that purpose.

[…]

The Copyright (Librarians and Archivists) (Copying of Copyright Material) Regulations

[…]

Interpretation

2. In these Regulations -

"the Act" means the Copyright, Designs and Patents Act 1988;

"the archivist" means the archivist of a prescribed archive;

"the librarian" means the librarian of a prescribed library;

"prescribed archive" means an archive of the descriptions specified in paragraph (4) of regulation 3 below;

"prescribed library" means a library of the descriptions specified in paragraphs (1), (2) and (3) of regulation 3 below.

Descriptions of libraries and archives

3.—(1) The descriptions of libraries specified in Part A of Schedule 1 to these Regulations are prescribed for the purposes of section 38 and 39 of the Act:

Provided that any library conducted for profit shall not be a prescribed library for the purposes of those sections.

(2) All libraries in the United Kingdom are prescribed for the purposes of sections 41, 42 and 43 of the Act as libraries the librarians of which may make and supply copies of any material to which those sections relate.

(3) Any library of a description specified in Part A of Schedule 1 to these Regulations which is not conducted for profit and any library of the description specified in Part B of that Schedule which is not conducted for profit are prescribed for the purposes of sections 41 and 42 of the Act as libraries for which copies of any material to which those sections relate may be made and supplied by the librarian of a prescribed library.

(4) All archives in the United Kingdom are prescribed for the purposes of sections 42 and 43 of the Act as archives which may make and supply copies of any material to which those sections relate and any archive within the United Kingdom which is not conducted for profit is prescribed for the purposes of section 42 of the Act as an archive for which copies of any material to which that section relates may be made and supplied by the archivist of a prescribed archive.

(5) In this regulation "conducted for profit", in relation to a library or archive, means a library or archive which is established or conducted for profit or which forms part of, or is administered by, a body established or conducted for profit.

[…]

Copying by librarian or archivist for the purposes of replacing items in a permanent collection

6. —(1) For the purposes of section 42 of the Act the conditions specified in paragraph (2) of this regulation are prescribed as the conditions which must be complied with before the

librarian or, as the case may be, the archivist makes a copy from any item in the permanent collection of the library or archive in order to preserve or replace that item in the permanent collection of that library or archive or in the permanent collection of another prescribed library or archive.

(2) The prescribed conditions are -

(a) that the item in question is an item in the part of the permanent collection maintained by the library or archive wholly or mainly for the purposes of reference on the premises of the library or archive, or is an item in the permanent collection of the library or archive which is available on loan only to other libraries or archives;

(b) that it is not reasonably practicable for the librarian or archivist to purchase a copy of that item to fulfil the purpose under section 42(1)(a) or (b) of the Act;

(c) that the other prescribed library or archive furnishes a written statement to the effect that the item has been lost, destroyed or damaged and that it is not reasonably practicable for it to purchase a copy of that item, and that if a copy is supplied it will only be used to fulfil the purpose under section 42(1)(b) of the Act; and

that the other prescribed library or archive shall be required to pay for the copy a sum not less than the cost (including a contribution to the general expenses of the library or archive) attributable to its production.

[…]

The Defamation Act 1996

s1. - (1) In defamation proceedings a person has a defence if he shows that-

> (a) he was not the author, editor or publisher of the statement complained of,
>
> (b) he took reasonable care in relation to its publication, and
>
> (c) he did not know, and had no reason to believe, that what he did caused or contributed to the publication of a defamatory statement.

(2) For this purpose "author", "editor" and "publisher" have the following meanings, […]

> "author" means the originator of the statement, but does not include a person who did not intend that his statement be published at all;
>
> "editor" means a person having editorial or equivalent responsibility for the content of the statement or the decision to publish it; and
>
> "publisher" means a commercial publisher, that is, a person whose business is issuing material to the public, or a section of the public, who issues material containing the statement in the course of that business.

(3) A person shall not be considered the author, editor or publisher of a statement if he is only involved-

> (a) in printing, producing, distributing or selling printed material containing the statement;
>
> [...]
>
> (c) in processing, making copies of, distributing or selling any electronic medium in or on which the statement is recorded, or in operating or providing any equipment, system or service by means of which the statement is retrieved, copied, distributed or made available in electronic form;

(4) Employees or agents of an author, editor or publisher are in the same position as their employer or principal to the extent that they are responsible for the content of the statement or the decision to publish it.

(5) In determining for the purposes of this section whether a person took reasonable care, or had reason to believe that what he did caused or contributed to the publication of a defamatory statement, regard shall be had to-

> (a) the extent of his responsibility for the content of the statement or the decision to publish it,
>
> (b) the nature or circumstances of the publication, and
>
> (c) the previous conduct or character of the author, editor or publisher.
>
> [...]

s2 (3) An offer to make amends-

> (a) must be in writing,

(b) must be expressed to be an offer to make amends under section 2 of the Defamation Act 1996, and

(c) must state whether it is a qualified offer and, if so, set out the defamatory meaning in relation to which it is made.

(4) An offer to make amends under this section is an offer-

(a) to make a suitable correction of the statement complained of and a sufficient apology to the aggrieved party,

(b) to publish the correction and apology in a manner that is reasonable and practicable in the circumstances, and

(c) to pay to the aggrieved party such compensation (if any), and such costs, as may be agreed or determined to be payable.

The Electronic Commerce (EC Directive) Regulations 2002

[…]

Mere conduit

17. - (1) Where an information society service is provided which consists of the transmission in a communication network of information provided by a recipient of the service or the provision of access to a communication network, the service provider (if he otherwise would) shall not be liable for damages or for any other pecuniary remedy or for any criminal sanction as a result of that transmission where the service provider -

(a) did not initiate the transmission;

(b) did not select the receiver of the transmission; and

(c) did not select or modify the information contained in the transmission.

(2) The acts of transmission and of provision of access referred to in paragraph (1) include the automatic, intermediate and transient storage of the information transmitted where:

(a) this takes place for the sole purpose of carrying out the transmission in the communication network, and

(b) the information is not stored for any period longer than is reasonably necessary for the transmission.

Caching

18. Where an information society service is provided which consists of the transmission in a communication network of information provided by a recipient of the service, the service provider (if he otherwise would) shall not be liable for damages or for any other pecuniary remedy or for any criminal sanction as a result of that transmission where -

(a) the information is the subject of automatic, intermediate and temporary storage where that storage is for the sole purpose of making more efficient onward transmission of the information to other recipients of the service upon their request, and

(b) the service provider -

(i) does not modify the information;

(ii) complies with conditions on access to the information;

(iii) complies with any rules regarding the updating of the information, specified in a manner widely recognised and used by industry;

(iv) does not interfere with the lawful use of technology, widely recognised and used by industry, to obtain data on the use of the information; and

(v) acts expeditiously to remove or to disable access to the information he has stored upon obtaining actual knowledge of the fact that the information at the initial source of the transmission has been removed from the network,

or access to it has been disabled, or that a court or an administrative authority has ordered such removal or disablement.

Hosting

19. Where an information society service is provided which consists of the storage of information provided by a recipient of the service, the service provider (if he otherwise would) shall not be liable for damages or for any other pecuniary remedy or for any criminal sanction as a result of that storage where -

(a) the service provider -

(i) does not have actual knowledge of unlawful activity or information and, where a claim for damages is made, is not aware of facts or circumstances from which it would have been apparent to the service provider that the activity or information was unlawful; or

(ii) upon obtaining such knowledge or awareness, acts expeditiously to remove or to disable access to the information, and

(b) the recipient of the service was not acting under the authority or the control of the service provider.

[…]

Defence in Criminal Proceedings: burden of proof

21. - (1) This regulation applies where a service provider charged with an offence in criminal proceedings arising out of any transmission, provision of access or storage falling within regulation 17, 18 or 19 relies on a defence under any of regulations 17, 18 and 19.

(2) Where evidence is adduced which is sufficient to raise an issue with respect to that defence, the court or jury shall assume that the defence is satisfied unless the prosecution proves beyond reasonable doubt that it is not.

Notice for the purposes of actual knowledge

22. In determining whether a service provider has actual knowledge for the purposes of regulations 18(b)(v) and 19(a)(i), a court shall take into account all matters which appear to it in the particular circumstances to be relevant and, among other things, shall have regard to -

(a) whether a service provider has received a notice through a means of contact made available in accordance with regulation 6(1)(c), and

(b) the extent to which any notice includes -

(i) the full name and address of the sender of the notice;

(ii) details of the location of the information in question; and

(iii) details of the unlawful nature of the activity or information in question.

The Data Protection Act 1998

[…]

Research, history and statistics.

s.33. - (1) In this section-

"research purposes" includes statistical or historical purposes;

"the relevant conditions", in relation to any processing of personal data, means the conditions-

> (a) that the data are not processed to support measures or decisions with respect to particular individuals, and
>
> (b) that 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.

(2) For the purposes of the second data protection principle, the further processing of personal data only for research purposes in compliance with the relevant conditions is not to be regarded as incompatible with the purposes for which they were obtained.

(3) Personal data which are processed only for research purposes in compliance with the relevant conditions may, notwithstanding the fifth data protection principle, be kept indefinitely.

(4) Personal data which are processed only for research purposes are exempt from section 7 if-

> (a) they are processed in compliance with the relevant conditions, and
>
> (b) the results of the research or any resulting statistics are not made available in a form which identifies data subjects or any of them.

(5) For the purposes of subsections (2) to (4) personal data are not to be treated as processed otherwise than for research purposes merely because the data are disclosed-

> (a) to any person, for research purposes only,
>
> (b) to the data subject or a person acting on his behalf,
>
> (c) at the request, or with the consent, of the data subject or a person acting on his behalf, or
>
> (d) in circumstances in which the person making the disclosure has reasonable grounds for believing that the disclosure falls within paragraph (a), (b) or (c).

Appendix B - License for Deposit of Web Materials

Licence Form for Deposit of Web Materials

1. Parties and Contact Details

(1) Printed name .....
(hereafter 'the Depositor')

Signed .....

Date .....

Position .....

Institution .....

Address .....

.....

Telephone ..... Fax .....

E-mail .....@.....

and

(2) Printed name .....
for the [web archive]

Signed .....

Date .....

Position .....

Address .....

.....

Telephone ..... Fax .....

E-mail .....@.....

2. Introduction

2.1 The depositor wishes to deposit material for archiving and distribution by the [web archive operator] *for education, private study, and research ("educational purposes") OR for public access OR for [specific purpose]*.

2.2 The [web archive operator] is funded by the Joint Information Systems Committee (JISC) and the Wellcome Trust to provide an archiving and distribution service for [content topic] web-based materials.

2.3 This agreement between the Depositor and the [web archive operator] provides the legal permissions and warranties needed to allow the [web archive operator] to preserve, and make accessible the deposited materials for *educational purposes OR for public access OR for [specific purpose]*.

2.4 This is a non-exclusive licence, which ensures that copyright in the original material is not transferred by this agreement and provides other safeguards for the depositor such as requesting acknowledgement in any publications arising from future research using the [material collected]. It permits use of the [web-based material] *only for non-commercial, teaching, research and private study OR for public access OR for [specific purpose]. Access to the [web-based material] will only be available to authorised users who have agreed to abide by licence conditions unless the depositor has stated that the [web-based material] can be available to any user.*

3. Definitions and Interpretation

3.1 In this Agreement the following words have the following meanings:

'Agreement'	this document including all of its terms and conditions and the Data and Documentation Transfer Form providing a schedule of the [material collected].
'Authorised user'	individuals authorised and registered by [web archive operator] to use the [archive name] or a member of an institution authorised and registered by [web archive operator] to use the [archive name] under a site licence.
'the [material collected]'	the material to be provided by the Depositor under the title in the Data and Documentation Transfer Form providing a schedule of the [material collected].
Commercial purposes	use of the [web-based material] for any reason direct or indirect which generates a profit
Educational purposes	use of the [web-based material] for education, private study or research provided that such use does not generate a profit.

4. Licence

4.1 The Depositor grants a non-exclusive licence of the [web-based material] to the [web archive operator] for the duration of this Agreement for archiving, distribution and use for *educational purposes OR for public access OR for [specific purpose]*. Such right shall include (but not be limited to) the right to:

4.1.1 issue copies of the [web-based material] to authorised users in a variety of media formats

4.1.2 promote and advertise the [web-based material] in any publicity for the [archive name], [web archive operator], JISC or Wellcome Trust.

4.1.3 to catalogue, enhance, validate and document the [material collected]

4.1.4 to electronically store, translate, copy, or re-arrange the [web-based material] to ensure its future preservation and accessibility

5. Depositor's rights and undertaking

5.1 The Depositor undertakes to make the [web-based material] available to the [web archive operator] as follows:

5.1.1 by permitting the [web archive operator] to download the [web-based material] from the Depositor's publicly accessible website, or

5.1.2 by providing the [web archive operator] with the necessary means of access to the Depositor's restricted access website or database to permit the downloading of the [web-based material], or

5.1.3 by providing the [web archive operator] with the [web-based material] in digital form on media agreed with the [web archive operator]

5.2 The Depositor does not warrant or guarantee the [web-based material] in terms of the comprehensiveness, accuracy, reliability, or otherwise of its contents.

5.3 The Depositor hereby warrants and undertakes as follows:

5.3.1 that the Depositor is the owner of the copyright and associated intellectual property rights in the whole [web-based material] or is duly authorised by the owner, or owners, of these rights and is capable of granting under this agreement, a licence to hold and disseminate copies of the material.
[See variants]

5.3.2 that the Depositor is the owner of any performance rights associated with all or part of the Work, or that where one or more persons, other than the Depositor, have performance rights associated with all or part of the Work to be deposited, the Depositor has either a written waiver of those rights, or permission in writing permitting the [web archive operator] to hold and disseminate copies of the material, or an agreement has been reached with the relevant licencing and collecting body or bodies, that the [web archive operator] may hold and disseminate copies of the material.
[See variants]

5.3.3 that the Work to be deposited was not made in breach of any exclusive recording right, and that where all or part of the Work is a recording of a performance in which a person or persons other than the Depositor has an exclusive recording right, a waiver of those rights or a licence granting to the [web archive operator] to hold and disseminate copies of the material has been obtained.

5.3.4 that the [web-based material] is not and shall be in no way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

5.3.5 that the [web-based material] does not and will not contravene any laws, including but not limited to, the law relating to defamation, or obscenity.

5.3.6 that the Depositor is not under any obligation or disability created by law contract or otherwise which would in any manner or to any extent prevent or restrict him from entering into and fully performing this Agreement.

5.3.7 to notify the [web archive operator] of any change of copyright ownership affecting the [material collected].

5.3.8 to notify [web archive operator] of any confidentiality, privacy or data protection issues pertaining to the [material collected]

6. The [web archive operator]'s Rights and Responsibilities

6.1 The [web archive operator] shall:

6.1.1 take reasonable measures to prevent unauthorised access to, duplication of, or distribution of, the [web-based material] whilst it is in the [web archive operator]'s possession or under its control

6.1.2 permit authorised users to access and use the [material collected], or any part of it. All subsequent access to and use of such material will be for the authorised user's educational purpose and may not be offered, whether for sale or not, to anyone who is not an authorised user.

6.1.3 draw the following notice to the attention of each authorised user as part of the authorisation process:

"All material supplied via the [web archive operator] is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the [archive name] is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. Permission for any other use must be obtained from the relevant copyright holder. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user."

6.1.4 request authorised users publishing any work based in whole or in part on the [web-based material] to display information crediting its creator and depositor.

6.1.5 not be under any obligation to take legal action on behalf of the Depositor or other rightsholders in the event of breach of intellectual property rights or any other right in the material deposited

6.1.6 not be under any obligation to reproduce, transmit, broadcast, or display the [web-based material] in the same formats or resolutions as those noted in the Data and Documentation Transfer Form.

6.2 While every care will be taken to preserve the physical integrity of the [material collected], the [web archive operator] shall incur no liability, either express or implicit, for the [web-based material] or for the loss of or damage to any of the [material collected].

6.3 The copyright in any additional data added by the [web archive operator] to the [material collected], and any search software, user guides and documentation that are prepared by [web archive operator]] to assist authorised users in using the [archive name] shall belong to [web archive operator] and any other parties that the [web

archive operator] may choose to enter into an agreement with to produce such materials.

7. Royalties

7.1 No royalties shall be paid for the use of the [web-based material] for educational purposes, archiving or publicity.

8. General

8.1 **Communications**
All notice under this Agreement shall be in writing and shall be sent to the address of the recipient set out in this Agreement or to such other address as the recipient may have notified from time to time. Any notice may be delivered personally or by first class post or by fax or by e-mail and shall be deemed to have been served if by hand when delivered, if by first class post 48 hours after posting, if by fax when confirmation of transmission is received and if by e-mail, when confirmation of receipt is received from the system of the recipient. If no reply is received to a notice under this agreement the consent of the recipient will be deemed to have been given after 30 days have elapsed from the issue of that notice.

8.2 **Successors**
This agreement is binding on and will benefit the successors and assigns of the parties.

8.3 **Entire Agreement**
This Agreement constitutes the entire agreement between the parties. No variation will be effective unless in writing signed by or on behalf of both parties.

8.4 **Invalidity**
If any part of this Agreement is held unlawful or unenforceable that part shall be struck out and the remainder of this Agreement shall remain in effect.

8.5 **Joint Venture**
This Agreement does not create any partnership or joint venture between the parties

8.6 **Waiver**
No delay neglect or forbearance by either party in enforcing its rights under this Agreement shall be a waiver of or prejudice of those rights

8.7 **Proper Law**
This Agreement is governed by the laws of England excluding any conflicts of law principles. Any dispute that may arise concerning this Agreement shall be decided by the High Court and the parties shall submit to its exclusive jurisdiction for that purpose.

8.8 **Term of the Agreement**
This Agreement shall take effect on execution hereof and shall continue for the duration of copyright in the [web-based material] unless either party terminates this agreement.

8.9 **Termination**

8.9.1 In addition to any remedy, the [web archive operator] on the one hand and the Depositor on the other may terminate this agreement immediately without further obligation in the event of any breach of this Agreement which cannot be remedied or

is not remedied within thirty (30) days of the party in breach being requested to do so by the other party.

8.9.2 Where there is no breach, either party may terminate this Agreement upon 6 months notice. However, where there is no breach and this Agreement is terminated by the Depositor during the term of the agreement the [web archive operator] shall be entitled to charge the Depositor for such costs as have been incurred in archiving and cataloguing the [material collected], and any other investment of resources in the [material collected], prior to its withdrawal.

8.10 **Disclaimer**
The Depositor and the [web archive operator] shall be under no liability for any loss or for any failure to perform any obligation hereunder due to causes beyond their control, including but not limited to industrial disputes of whatever nature, Acts of God, hostilities, force majeure or any circumstances which they could not reasonably foresee and provide against.

Variant Clauses

Variant for the depositor to waive some or all rights in the [material collected]

4. Licence

4.1 The Depositor wishes the [web-based material] to be freely available to any user and for any purpose. To this end, the Depositor {waives all copyrights in the [web-based material] for their full duration in all jurisdictions/assigns all copyrights in the [web-based material] for their full duration in all jurisdictions to the [web archive operator] }.

4.2 The depositor {also waives/does not waive} the following rights as applicable:

4.2.1 the moral right of paternity (the right to be identified as the author or director of the work), and

4.2.2 the moral right to object to derogatory treatment of a copyright work.

4.2.3 any performance right that they may have in the work.

Variants for where the Depositor is not the owner of the copyright or associated intellectual property rights

5.3 The Depositor hereby warrants undertakes and agrees with the [web archive operator] as follows:

5.3.1 that where the Depositor is not the owner of the copyright and associated intellectual property rights in the [material collected], or any part thereof, or is not duly authorised by the owner(s) of those rights, that a Copyright Licence form has been completed by the owner(s) of the copyright and associated intellectual property rights in the whole or relevant part(s), granting to the [web archive operator] a licence to hold and disseminate copies of the material.

or

that where the Depositor is not the owner of the copyright and associated intellectual property rights in the [web-based material] or any part thereof, that an agreement has been reached with the relevant licencing and collecting body or bodies, that a licence can be granted to the [web archive operator] to hold and disseminate copies of the material.

or

that where the Depositor is not the owner of the copyright and associated intellectual property rights in the [web-based material] nor duly authorised by the owner(s) of the copyright and associated intellectual property rights, the Depositor has;

5.2.1.1 made all reasonable inquiries to ascertain the identity of the author(s) of the relevant material, but failed to so ascertain, and

5.2.1.2 has reason to assume that the copyright has expired or that the author(s) died 70 years or more before the beginning of the year in which this Agreement comes into force.}

Variant covering performance rights

Performance Rights

The Depositor hereby warrants undertakes and agrees with the [web archive operator] as follows:

5.32 that where one or more persons, other than the Depositor, have performance rights associated with all or part of the [web-based material] to be deposited, the Depositor has a written waiver of those rights.

or

that where one or more persons, other than the Depositor, have performance rights associated with all or part of the [web-based material] to be deposited, the Depositor has permission in writing from those persons granting to the [web archive operator] a licence to hold and disseminate copies of the material.

or

that where one or more persons, other than the Depositor, have performance rights associated with all or part of the Work to be deposited, an agreement has been reached with the relevant licencing and collecting body or bodies, that a licence can be granted the [web archive operator] to hold and disseminate copies of the material.

and/or

that the [web-based material] to be deposited was not made in breach of any exclusive recording right, and that where all or part of the [web-based material] is a recording of a performance in which a person or persons other than the Depositor has an exclusive recording right, a waiver of those rights or a licence granting to the [web archive operator] permission to hold and disseminate copies of the material has been obtained.

Variant for online access and registration

The [web archive operator] shall:

6.1.X Register a user as an authorised user of the Depositor's [web-based material] only after receipt of a signed Common Access Agreement Form, or where the user has signed an equivalent institutional document as required under the Site Licence Agreement, when notification of that signature is received by the [web archive operator]. In this clause, 'signed' means the hand-written signature of the user; it does not mean any form of 'electronic signature'

Or

6.1.X Register a user on-line as an authorised user of the Depositor's [web-based material] only after the Common Access Agreement has been displayed to the user, and the user has completed an on-line form providing [*details*] [subject to verification by [web archive operator]] and has agreed to abide by those terms and conditions.

Or

6.1.X Register a user on-line as an authorised user of the Depositor's [web-based material] only after the Common Access Agreement has been displayed to the user, and the user has indicated his intent to abide by those terms and conditions by [clicking on an "Accept" button" etc.].

Variants for charging

New Section **Charges**

X.X The Depositor will pay [web archive operator] an annual fee to cover administrative, processing, and archiving costs for the [material collected]. The Depositor will be provided with an Annual Fee Schedule detailing fee and payment date on [date] each year.

X.X Failure to pay the requisite annual fee will result in [web archive operator] being unable to continue to support the archiving of the [material collected], and, where reasonable, will result in its return to the Depositor, upon payment of any outstanding administrative, processing, and archiving costs incurred by [web archive operator].

X.X Deposit charges and annual fees may be waived by [web archive operator] either permanently, or for the duration of the funding, where core funding has been obtained for [web archive operator] archiving of the [web-based material] from a granting agency or other body.

Or

X.X The Depositor will pay [web archive operator] a fee of [£???] prior to deposit of the [web-based material] to cover administrative, processing, and archiving costs.

And

X.X All fees paid by the Depositor to [web archive operator] to cover administrative, processing, and archiving costs for the [web-based material] are non-refundable.