



department for
**culture, media
and sport**

Proposal on the Collection and Preservation of UK Offline and Microform Publications and UK Online Publications

Available free of charge and without access restrictions

Legal Deposit Libraries Act 2003
December 2009

improving
the quality
of life for all

Our aim is to improve the quality of life for all through cultural and sporting activities, support the pursuit of excellence, and champion the tourism, creative and leisure industries.

Contents

Chapter 1: General Information	4
Chapter 2: Purpose	5
Annex A: Proposals for Offline and Microform Publications	7
Annex B: Proposals for Online Publications	9
Annex C: Online Content to be Published	12
Annex D: Impact Assessment – UK Online Publications, which are available free of charge and without access restrictions	17
Annex E: Further Details on Territoriality	35
Annex F: Further Details on Harvesting Process	36
Annex G: Questionnaire	38

Chapter 1: General Information

- 1.1. The purpose of legal deposit is to ensure that the nation's published output (and thereby its intellectual record and future published heritage) is collected systematically, and as comprehensively as possible, both in order to preserve the material for the use of future generations and, with certain important exceptions (set out later in this document), to make it available to readers within the libraries¹ defined in the Legal Deposit Libraries Act 2003 (2003 Act). The system dates back several hundred years, enforced by statute since 1662, and has been a vital element in preserving and making available the published record of previous generations for readers today and the in future.
- 1.2. The 2003 Act² reaffirmed existing provisions for deposit of printed publications and created a framework of legislation in which regulations could be made for the deposit of electronic and other non-print publications. The Legal Deposit Advisory Panel (LDAP), an independent non-departmental public body, was set up in September 2005 to work at arm's length from the Government and to advise them on the implementation of the Act and to make recommendations on regulatory options for the deposit of non-print publications.
- 1.3. These proposals cover the deposit of United Kingdom non-print publications in microform and offline electronic media and UK Online Publications, which are available free of charge and without access restrictions.
- 1.4. A separate proposal will be produced on UK Commercial and Protected Online Publications following further recommendations from LDAP.

¹ The British Library; the National Library of Scotland; the National Library of Wales; the Bodleian Library, Oxford; the University Library, Cambridge; and the Library of Trinity College, Dublin

² = See: http://www.opsi.gov.uk/acts/acts2003/ukpga_20030028_en_1

Chapter 2: Purpose

- 2.1 The purpose of this document is to set out for consultation the recommendations provided by LDAP to the Secretary of State for DCMS on legal deposit of non-print material – specifically UK Offline and Microforms Publications (Annex A) and UK Online Publications (Annex B to F), which are available free of charge and without access restrictions with the Legal Deposit Libraries³ (the Libraries).
- 2.2 We are seeking your views on all aspects of the recommendations and have also set some specific questions to which we would welcome your answers (Annex G).
- 2.3 Following this consultation, and further consultation on the detail of the regulations as appropriate, our aim is to introduce affirmative secondary legislation to Parliament that will lead to regulations for publishers of relevant non-print publications to deposit one or more copies with the Libraries, as detailed in the following provisions.
- 2.4 The consultation period will run for 12 weeks from 8 December to 1 March 2010.
- 2.5 Please respond before the closing date, using the questionnaire at Annex G, to deposits.consultation@culture.gsi.gov.uk. If you do not have access to email, please write to:
- Frances Love**
Libraries and Archives Team
Culture Directorate
2-4 Cockspur Street
London SW1Y 5DH
- 2.6. For enquiries about the consultation (handling) process only please contact the DCMS Public Engagement and Recognition Unit (PERU) at the above address or email using the form at www.culture.gov.uk/contact_us heading your communication 'Proposal on the Collection and Preservation of UK Offline and Microform Publications and UK Online Publications, which are available free of charge and without access restrictions'.
- 2.7. Copies of responses will be published after the consultation closing date on the Department's website: www.culture.gov.uk
- 2.8 Information provided in response to this consultation, including personal information, may be published or disclosed in accordance with the access to information regimes (primarily the Freedom of Information Act 2000 (FOIA), the Data Protection Act 1998 (DPA) and the Environmental Information Regulations 2004). If you want the information that you provide to be treated as confidential, please be aware that, under the FOIA, there is a statutory Code of Practice with which public authorities must comply and which deals, amongst other things, with obligations of confidence. In view of this it would be helpful if you could explain to us why you regard the information you have provided as confidential. If we receive a request for disclosure of the information we will take full account of your explanation, but we cannot give an assurance that confidentiality can be maintained in all

³ The British Library; the National Library of Scotland; the National Library of Wales; the Bodleian Library, Oxford; the University Library, Cambridge; and the Library of Trinity College, Dublin.

circumstances. An automatic confidentiality disclaimer generated by your IT system will not, of itself, be regarded as binding on the Department.

- 2.9 The Department will process your personal data in accordance with the DPA, and in the majority of circumstances, this will mean that your personal data will not be disclosed to third parties.
- 2.10 The consultation is guided by the Government's Code of Practice on Consultation which is available at: <http://www.berr.gov.uk/whatwedo/bre/code/page46954.html>

Annex A: Proposals for Offline and Microform Publications

Introduction

- 3.1 LDAP was set up in September 2005 to advise the Secretary of State on the implementation of the 2003 Act and to make recommendations. They have worked at arm's length from Government and developed their own work programme concentrating on three areas of non-print material for the collection and preservation of:
- UK Offline and Microforms Publications;
 - UK Online Publications, free of charge and without access restrictions;
 - UK Commercial and Protected Online Publications.
- 3.2 This Chapter looks at the recommendations LDAP made for Offline and Microform Publications.

LDAP's Recommendations

- 3.3 For Offline and Microform Publications LDAP considered four options:
- Option 1: Publisher Archiving
 - Option 2: Unregulated Voluntary Code of Practice, no active requesting by libraries
 - Option 3: Self-Regulated Voluntary Scheme, with active requesting by libraries (recommended)
 - Option 4: Statutory regulations.

Option 1

- 3.4 **Publisher Archiving.** Under this option the libraries would not seek to build a central archive, and the nation would rely entirely upon the private sector. Publishers would retain copies of their own publications specifically for archival purposes, with various access and cost arrangements, resulting in several archives behind multiple commercial barriers (this is not entirely dissimilar to the current publisher practice of maintaining back catalogues or online archives).

Option 2

- 3.5 **Unregulated Voluntary Code of Practice.** In 2000, representatives from the Legal Deposit Libraries and the publishing Trade Associations drew up a Code of Practice for the Voluntary Deposit of Microform and Offline Publications. This encouraged publishers to deposit with Libraries; however, they could choose whether and how to deposit. Under this option the Code would become permanent, encouraging publishers to deposit offline and microform publications. However, there would be no governance or additional support from industry bodies apart from publicising the Code itself. In addition, the Libraries would not actively identify or request publications.

Option 3

- 3.6 **Self-Regulated Voluntary Scheme, with active requesting by Libraries.** This option would enhance the Voluntary Code above by establishing a formal, lasting agreement and governance structure via LDAP, to include Memoranda of Understanding signed by Libraries and publishers, a monitoring and annual reporting scheme, and five-yearly reviews both of effectiveness of the scheme and of the trend in the number of relevant publications. Publishers would be actively encouraged to deposit by their Trade Association representatives; Libraries would actively identify relevant publications and request deposit when necessary.

Option 4

- 3.7 **Statutory regulations.** Government regulation, under the Legal Deposit Libraries Act 2003, would oblige publishers to deposit, free of charge, one copy with the British Library and, upon request, up to five additional copies for the remaining libraries, on the same basis as already applies to printed publications. The Libraries would be entitled (but not legally obliged) to identify relevant publications that had not already been deposited and to claim them from the publishers.

Government Proposals

- 3.8 We have considered the detailed proposals provided by LDAP and are content to accept their recommendation of a **Self-Regulated Voluntary Scheme, with active requesting by Libraries.**
- 3.9 We reserve the right to monitor the scheme to ensure that the nation's published output (and thereby its intellectual record and future published heritage) is collected systematically, and as comprehensively as possible. Should the need arise we will regulate at the appropriate time under the powers available to us in the 2003 Act.
- 3.10 Therefore we are not seeking any formal responses on this proposal.

Annex B: Proposals for Online Publications

Introduction

- 4.1 These proposals cover UK Online Publications which are available free of charge and without access restrictions.
- 4.2 They **do not** include UK Commercial and Protected Online Publications.

LDAP's Recommendations

- 4.3 For online publication LDAP considered three options:
- Option 1: Legal Deposit Library Permissions-based Harvesting and Archiving
 - Option 2: Legal Deposit Library regulation-based Harvesting and Archiving (Recommended)
 - Option 3: Archiving Left to the Market

Option 1

- 4.4 **Legal Deposit Libraries Permissions-based Harvesting and Archiving.** This option might be considered as the equivalent of a voluntary code for deposit, in that it will not be subject to regulation, but it presumes that legal deposit Libraries will actively harvest ('pull') material, rather than imposing any duty or encouragement upon publishers to deposit. Thus the main burden of activity rests on the Libraries to take periodic copies of websites and documents, then ingest them into an archive for preservation. However Libraries may only do so after obtaining copyright permission in writing from the publishers or owners of the material; in responding to these requests, online publishers must give permission on behalf of third party rights-holders.

Option 2

- 4.5 **Legal Deposit Libraries Regulation-based Harvesting and Archiving.** This option presumes that libraries, or agents on their behalf, will actively harvest ('pull') material, rather than imposing any duty or encouragement upon publishers to deposit. Thus the main burden of activity rests on the Libraries to take periodic copies of freely available online publications and documents, then ingest them into an archive. However regulation would allow Libraries to do so without needing to contact the publishers or owners of the material.

Option 3

- 4.6 **Archiving Left to the Market.** This option proposes there is no central, public sector initiative, and that archiving would be left to the private sector and market forces. Moreover, Libraries would not actively seek to harvest or build a national archive for freely available online publications. Individual publishers might approach a Library to archive material on its behalf; this would be subject to separate negotiations and cost-benefit analysis for the specific circumstances.

Question 1: What are your views on the options considered for this content?

Question 2: Are there any other options that should have been considered? If so what are they?

LDAP'S Analysis

- 4.7 Permissions-based harvesting (Option One) would capture a minute proportion of freely available online publications. In addition, a less than comprehensive collection would come at a much higher cost per unit for publishers and Libraries. Permission is only successfully obtained in 30% of cases (based on UKWAC figures).
- 4.8 Archiving left to the market (Option Three) makes it impossible to deliver a comprehensive or freely available collection. Whilst some professional publishers do archive their own material, all the available evidence suggests that the vast majority of individual publishers of freely available online publications within the UK domain do not. Therefore a substantial proportion of the UK's cultural and intellectual output will be lost forever.
- 4.9 LDAP recommends Regulation-based Harvesting and Archiving (Option Two) for the deposit of freely available online publications, to enable Libraries to gather and preserve publications as defined within the Recommendation, and make them available according to the 2003 Act. This deposit option secures the benefits inherent in preserving for future generations a comprehensive archive of the UK's online intellectual and cultural heritage (it is estimated that up to 79% of material in scope could be captured and archived within available resources), at a manageable and proportionate cost, not only for publishers and libraries, but also to the public.

Question 3: Do you agree with the analysis of these options? Explain why.

- 4.10 **Rate of Deposit:** The large number of freely available online publications and the rate of increase (about 15% per annum) expected pose a challenge to creating a comprehensive collection in the most efficient way possible. Harvesting provides the most efficient and timely solution for deposit of publications in this category: regulations-based harvesting specifically ensures the greatest amount can be collected to build the most comprehensive collection for future generations.

Question 4: Do you agree that harvesting provides the most efficient and timely solution for deposit of publications in this category? Explain why.

- 4.11 **Proportionality:** Regulation is an appropriate response to the enormous number of publishers and publications in this category. It is the most cost-efficient method of collection for the Libraries and imposes no direct financial or administrative burden upon the publishers.

Question 5: Do you agree regulation is the most cost-efficient method of collection for the Libraries and imposes no direct financial or administrative burden upon the publishers? Explain why.

- 4.12 **Existing Harvesting Activities and Infrastructure:** Whilst Libraries have not yet been able to harvest widely across the whole UK domain, much of the technical infrastructure and expertise that they would need is already in place. They have practical experience of permissions-based harvesting for the UK Web Archiving Consortium. In addition, the International Internet Preservation Consortium (IIPC), of which the British Library and National Library Scotland (NLS) are members, is developing tools, infrastructure and the policies for addressing the challenges of collection and preservation

internationally. Moreover, knowledge gained from other national Libraries that already carry out domain-level harvesting (e.g. in France, Denmark, New Zealand and Australia) will also assist its implementation in the UK. Indeed, the UK legal deposit Libraries estimate that they could begin harvesting on this basis, up to 80% in scale, within existing resources and within a very short time of implementation.

- 4.13 **Civil Liabilities:** Although it cannot eliminate all potential risks, regulation would protect both publishers and Libraries from such civil liabilities as copyright infringement and defamation.

Annex C: Online Content to be Published

Defining the UK Web: Publications in Scope

- 5.1 The Recommendation only relates to freely available online publications, which can be harvested or collected by Legal Deposit Libraries (LDLs)⁴ without any requirement for action by publishers (a reflection that the publications are available to the public free of charge and accessible without restriction). Restrictions that would remove publications from the scope of these proposals may include identification, authentication/authorisation, registration, subscription, and Internet Protocol (IP) address range⁵. Material that requires compliance with a basic technical formality such as downloading 'cookies' should be permitted, provided that this does not entail any active (human) intervention by the publisher or website owner.
- 5.2 The online publications to which this Recommendation applies are not intended to include:
- Sites outside the UK (see Territoriality below)
 - Chargeable content/commercial content
 - Sites with technical barriers
 - Secured transactions
 - Members-only areas within public sites
 - Private intranets and restricted access content
 - Recorded sound and film where such works comprise the sole or main purpose of the content or where any other material is incidental (e.g. the BBC 'Radio Player', any equivalents of Napster, YouTube and suchlike, and sites offering ring tones or streamed films and programmes from broadcasters).

Question 6: Do you agree that this is an appropriate definition for the type of publications that should be included in scope for regulations? Explain why. Is there anything else that should be included in this definition? Is there anything that should be excluded from this definition?

Defining the UK Territoriality

- 5.3 Harvesting, where online publications are collected using software that facilitates their collection and archiving, provides a simple approach to deposit of such a wide range and number of publications. The first step in harvesting is defining the parameters for collection and its required links to the UK. This definition is also a requirement of Section 1 of the 2003 Act.

⁴For the purposes of this paper, 'LDLs' or 'Libraries' applies to all six Legal Deposit Libraries named in this Recommendation paper: the British Library, the National Library of Scotland, the National Library of Wales, and the University Libraries of Cambridge, Oxford, and Trinity College, Dublin.

⁵Where access is only enabled for users within a specified IP address range.

- 5.4 In fulfilment of these proposals, the territoriality criteria proposed for publications relevant to this Recommendation are that:
- Publishers should be based in the UK or have a UK address (physical or electronic);
 - Publications should be lawfully published or made available by or on behalf of that publisher from a UK address; and
 - Publications should be made available to the public.

The following criteria are also thought to be relevant to fixing the place of publication and, therefore, the potential relevance to any approved harvesting for the purposes of legal deposit, namely that the site from which the publication is harvested:

- has a UK domain name
 - relates to UK-based individuals or organisations which use other domain names, such as .org, .com, .net etc. or alternatives; and
 - can be demonstrated, if an overseas publication, to be made available by a UK-based publisher.
- 5.5 Exceptions to this definition include publications:
- with no connection to the UK⁶; and
 - substantially consisting of sound recording or films (see Act s1 (5) a, b).
- 5.6 While territoriality establishes the parameters of the domain to be harvested, analysis of the domain growth and size identifies the scale of the work and some of the key assumptions underlying the calculation of costs.
- 5.7 Further information on this issue is set out in Annex E.

Question 7: Do you agree with the territorial definition of the UK web? Explain why. Is there anything else that should be included in this definition? Is there anything that should be excluded from this definition?

The UK Domain⁷

- 5.8 The category definition and territoriality rules govern what may be collected. Within these, the model used to calculate costs for harvesting assumes that the UK web space is defined as all .UK domains registered by Nominet (6.1 million in mid-2007) plus approximately 50,000 other domains which can be readily identified as published in the UK. See Libraries key costs assumptions for cost model and further information on assumptions (Annex D).
- 5.9 It is estimated that the numbers will continue to grow by 17% per annum until 2011, then by 15% until 2016. However, 35% of the domains are inactive, i.e. registered but not live, or where static content can be 'de-duplicated' after a first harvest. A further 25% are primarily 'deep web' or protected publications outside the scope of this category. Overall, the number of online publications in scope is therefore estimated at 3.9 million in 2007 rising to 14.6 million in 2016.

Question 8: Do you agree with this analysis of the UK Web Domain? Explain why. What do you think the impact of your analysis would be?

⁶ LDAP is reviewing the use of this phrase in connection with online publications, as its inclusion here would imply that publications on non-UK related subjects, but by British authors, would be excluded from the archive. A number of agencies with helpful practices might also aid the LDLs in identifying publications 'connected to the UK', such as Internet Watch and Nominet.

⁷ See notes under Libraries Key Costs Assumptions, for sources used to support the assumptions for the growth and size of the domain.

Size of the Domain

- 5.10 The average size of websites (and therefore the number of copyright works and publications that they contain) has been growing significantly each year. However, the cost model assumes that most audio-visual content, one of the major causes of growth, is out of scope, and therefore a more modest 5% growth per annum is appropriate. The average size also varies dramatically, from circa five megabytes for 80% of sites to one gigabyte for 0.5% of sites; this model assumes a weighted average of 25 megabytes.

Harvesting the Web⁸

- 5.11 The proposed method of collecting and preserving such a large number of publications is to 'pull' (harvest) them from the Web. Harvesting is an automated process, where, through the use of special software, libraries can collect publications with no action required by publishers. The costs, impact evidence, and success rate for this type of harvesting are based on a pilot implemented by the UK Web Archiving Consortium (UKWAC). The pilot, commencing in 2003 for two years (extended to September 2007), involved the selection of freely available online publications to be preserved and archived. The Consortium has so far archived more than 2,700 publications and over 10,000 instances (see www.webarchive.org.uk).
- 5.12 Harvesting conducted as part of a regulation does not require the individual permissions of publishers, as exemptions from such liabilities as copyright infringement and defamation are covered under the 2003 Act.

Question 9: How do you see a Deposit Library driven system of web harvesting interfacing with a publisher driven duty to deposit under the 2003 Act?

Question 10: How could Deposit Libraries most efficaciously ensure a comprehensive body of eligible content is deposited?

UK Legal Deposit Libraries Harvesting the Web

Harvesting Costs

- 5.13 This proposal involves harvesting by Libraries, therefore, the costs largely accrue to them. However, this does not impose a specific duty upon libraries to collect a pre-determined number or proportion of UK publications. Their duty is to collect in accordance with their overriding legal deposit obligations, to archive as much of the national cultural record and make it available for research within limitations of their resources and budgets. Therefore, these costs are not direct, bottom-line (cause and effect) consequences of each option. They are illustrations of what the libraries believe might realistically be achieved within their budget and resource constraints and after prioritising this activity and category of publications against other collection goals.
- 5.14 The cost of storage includes built-in redundancy to ensure safe preservation of the archive. However, the real cost of storage per terabyte has fallen by more than 30% per annum over the last 20 years and is expected to continue falling by 25% per annum until 2016.
- 5.15 Two infrastructures have already been designed and built (apart from certain elements) by the Libraries and will be used to store other digital or digitised collections as well as legal deposit material. Therefore, this cost model focuses only on the incremental systems costs (including renewing equipment every three years) plus staffing costs required to collect and preserve this category of publication.

⁸ Some stakeholders have reservations about the extent of harvesting and access to the harvested content and we will look at ways to overcome these concerns in our detailed policy proposals.

- 5.16 Costs have been analysed under the headings of selection, obtaining copyright permission, harvesting, QA, storage & preservation, resource discovery, digital rights management (DRM) & access, and other costs. They include salaries, pensions, NI and other staff-related costs, allocations for wider costs such as IT support and expenses, plus allocations for general overheads (See Annex D for more detail on costs and assumptions).
- 5.17 The near elimination of selection and IPR permissions activities makes harvesting a much more efficient process than requiring every publisher to deposit their own material. Total costs are estimated at £215 per annum for every terabyte archived over a 10-year period, although higher overall costs estimated at £1,132,000 per annum would be necessary for the infrastructure, harvesting, and storage, because of the greater volume collected⁹. See Libraries' key cost assumptions Annex D.
- 5.18 Further information on the practical arrangements are set out in Annex F.

Question 11: Do you agree with this costing model? Explain why. Are there costs that need to be factored in or excluded?

Publishers

- 5.19 Ascertaining publisher costs presents a difficulty that can be ascribed primarily to the broad definition of 'publisher' for this category of publication, a definition that is quite distinct from that of other categories for deposit. Traditionally, publishers are a group well defined and contained by type and content of publication, as well as by business model. The online publisher of freely available publications, however, runs the gamut from the individual blogger with no revenue stream to a multi-national corporation. This sheer number of publications and range of publisher types impose a considerable challenge for determining costs and benefits that suit any group of publishers, let alone cover the whole spectrum.
- 5.20 At the beginning of 2008, the Legal Deposit Advisory Panel undertook a survey of Trade Association publisher members, as well as non-commercial publishers that participated in the UKWAC pilot. This survey provided publishers with information about deposit as well as asked them for feedback on costs and other impacts of harvesting and archiving. The findings from the survey were as follows:
- A majority of those commercial and non-commercial publishers surveyed supported regulation-based harvesting;
 - Not only did they think this kind of harvesting the most efficient and less invasive to their business process, but they also observed that there would be relatively little cost to them;
 - However, publishers were not able to assess the level of cost to them associated with permissions-based harvesting.
- 5.21 Generally, publishers cost concerns were primarily in the area of revenue and the possible impact from harvesting, and to what extent these concerns could be addressed in a rapidly changing commercial and technological environment.
- 5.22 As publishers do not push (deposit) publications to libraries in the traditional sense, there appears to be no specific activity from which costs can be calculated. However, there are potential risks that may have significant impacts, if not eventual costs. These include copyright protection of freely available online publications. We are awaiting the outcomes from the UK Intellectual Property Office's Copyright Exceptions Consultation, so that this concern can be addressed in future detailed policy proposals.

⁹This figure represents the total cost across Legal Deposit Libraries. It assumes that readers in any of the six legal deposit libraries' premises would be able to access all materials and electronic publications that are harvested and archived by the BL/NLW/NLS infrastructures. The University Libraries of Oxford, Cambridge and Trinity College do not currently plan to harvest themselves to the same extent, but would retain the entitlement to do so.

- 5.23 Moreover, the deposit process adds a level of complexity for publishers in their agreements with third parties, either providing content or software. Indeed, there are concerns, as expressed in the Commercial Publishers Survey, over securing ongoing rights for data or images that were made available free of charge but on a time limited basis. For example, some promotional sites provide high value business information on a time-limited basis as sample data to encourage site traffic or subscription sales. Accordingly, publishers may be exposed to such liabilities as third party IPR and licensing infringement, as well as defamation, contempt of court, and libel.

Question 12: Do these assumptions adequately reflect the financial burden of publishers? Is there anything that needs to be included or excluded?

Annex D: Impact Assessment – UK Online Publications, which are available free of charge and without access restrictions

Summary: Intervention & Options		
Department /Agency: Department for Culture Media and Sport	Title: Impact Assessment of UK Online Publications, which are available free of charge and without access restrictions	
Stage: Consultation	Version: 1.0	Date: 30 November 2009
Related Publications: NA		
Available to view or download at: http://www.culture.gov.uk		
Contact for enquiries: www.culture.gov.uk/contact_us		Telephone:
<p>What is the problem under consideration? Why is government intervention necessary?</p> <p>The purpose of legal deposit is to ensure that the nation's published output (and thereby its intellectual record and future published heritage) is collected systematically, and as comprehensively as possible, both in order to preserve the material for the use of future generations. These proposals cover the deposit of United Kingdom Online Publications, which are available free of charge and without access restrictions; there is currently no system in place to allow for the collection of this content.</p>		
<p>What are the policy objectives and the intended effects?</p> <p>Following this consultation, and further consultation on the detail of the regulations as appropriate, our aim is to introduce affirmative secondary legislation to Parliament that will lead to regulations for publishers of relevant non-print publications to deposit one or more copies with the Libraries, as detailed in the following provisions. The objective is to preservice the material for future generation.</p>		
<p>What policy options have been considered? Please justify any preferred option.</p> <p>LDAP were set up in September 2005 to advise the Secretary of State on the implementation of the 2003 Act and to make recommendations. They have worked at arm's length from Government and have produced these recommendations. They considered three options:</p> <ol style="list-style-type: none"> 1. Permission-based Harvesting 2. Regulation-based Harvesting 3. Archiving Left to the Market 		

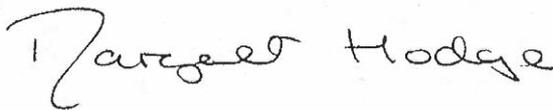
Summary: Intervention & Options (cont.)

When will the policy be reviewed to establish the actual costs and benefits and the achievement of the desired effects? A full impact assessment will be done before regulations are made.

Ministerial Sign-off For Consultation Stage Impact Assessments:

I have read the Impact Assessment and I am satisfied that, given the available evidence, it represents a reasonable view of the likely costs, benefits and impact of the leading options.

Signed by the responsible Minister:

A handwritten signature in black ink that reads "Margaret Hodge". The signature is written in a cursive style.

Date: 04/12/09

Summary: Analysis & Evidence

Policy Option: 1	Description: LDL Permissions-based Harvesting and Archiving
------------------	-------------------------------------------------------------

COSTS	ANNUAL COSTS		Description and scale of key monetised costs by 'main affected groups' Average total cost per annum to the Legal Deposit Libraries. These include technology, staffing, access and preservation costs.
	One-off (Transition)	Yrs	
	£		
	Average Annual Cost (excluding one-off)		
	£496,959		Total Cost (PV) £4.13m
Other key non-monetised costs by 'main affected groups'			

BENEFITS	ANNUAL BENEFITS		Description and scale of key monetised benefits by 'main affected groups'
	One-off	Yrs	
	£		
	Average Annual Benefit (excluding one-off)		
	£		Total Benefit (PV) £
Other key non-monetised benefits by 'main affected groups' There will be some preservation of the nation's published heritage, but not in a systematic manner. It is estimated that this option will only allow up to 0.5% of material included in scope to be captured.			

Key Assumptions/Sensitivities/Risks

Price Base Year 2009	Time Period Years 10	Net Benefit Range (NPV) £-4.13m	NET BENEFIT (NPV Best estimate) £-4.13m
--------------------------------	--------------------------------	-------------------------------------------	---------------------------------------------------

Summary: Analysis & Evidence (cont.)

What is the geographic coverage of the policy/option?	UK and Ireland			
On what date will the policy be implemented?	TBC			
Which organisation(s) will enforce the policy?	British Library			
What is the total annual cost of enforcement for these organisations?	£ NA			
Does enforcement comply with Hampton principles?	Yes			
Will implementation go beyond minimum EU requirements?	Yes			
What is the value of the proposed offsetting measure per year?	£			
What is the value of changes in greenhouse gas emissions?	£			
Will the proposal have a significant impact on competition?	No			
Annual cost (£-£) per organisation (excluding one-off)	Micro	Small	Medium	Large
Are any of these organisations exempt?	No	No	N/A	N/A

Impact on Admin Burdens Baseline (2005 Prices)			(Increase – Decrease)	
Increase of £	Decrease of £	Net Impact £		

Key:

Annual costs and benefits

Constant price

Summary: Analysis & Evidence

Policy Option: 2	Description: Legal Deposit Library Regulation-based Harvesting and Archiving
-------------------------	-------------------------------------------------------------------------------------

COSTS	ANNUAL COSTS		Description and scale of key monetised costs by 'main affected groups'. Average total cost per annum to the Legal Deposit Libraries. These include technology, staffing, access and preservation costs.
	One-off (Transition)	Yrs	
	£		
	Average Annual Cost (excluding one-off)		
	£1,132,286		Total Cost (PV) £9.42m
Other key non-monetised costs by 'main affected groups'			

BENEFITS	ANNUAL BENEFITS		Description and scale of key monetised benefits by 'main affected groups'
	One-off	Yrs	
	£		
	Average Annual Benefit (excluding one-off)		
	£		Total Benefit (PV) £
Other key non-monetised benefits by 'main affected groups' To ensure that the nation's published output (and thereby its intellectual record and future published heritage) is collected systematically, and as comprehensively as possible, both in order to preserve the material for the use of future generations. It is estimated that this option will allow up to 79% of material in scope could be captured and archived within available resources.			

Key Assumptions/Sensitivities/Risks

Price Base Year 2009	Time Period Years 10	Net Benefit Range (NPV) £-9.42m	NET BENEFIT (NPV Best estimate) £-9.42m
--------------------------------	--------------------------------	--------------------------------------------------	----------------------------------------------------------

Summary: Analysis & Evidence (cont.)

What is the geographic coverage of the policy/option?	UK and Ireland			
On what date will the policy be implemented?	TBC			
Which organisation(s) will enforce the policy?	British Library			
What is the total annual cost of enforcement for these organisations?	£ NA			
Does enforcement comply with Hampton principles?	Yes			
Will implementation go beyond minimum EU requirements?	Yes			
What is the value of the proposed offsetting measure per year?	£			
What is the value of changes in greenhouse gas emissions?	£			
Will the proposal have a significant impact on competition?	No			
Annual cost (£-£) per organisation (excluding one-off)	Micro	Small	Medium	Large
Are any of these organisations exempt?	No	No	N/A	N/A

Impact on Admin Burdens Baseline (2005 Prices)			(Increase – Decrease)	
Increase of £	Decrease of £	Net Impact £		

Key:

Annual costs and benefits

(Net) Present

Summary: Analysis & Evidence (cont.)

Policy Option: 3	Archiving Left to the Market
------------------	------------------------------

COSTS	ANNUAL COSTS		Description and scale of key monetised costs by 'main affected groups' There is no costs to the Libraries as it is purely up to the publishers to archive their material. There will be no obligation on them to do so, so cost can not be calculated.	
	One-off (Transition)	Yrs		
	£			
	Average Annual Cost (excluding one-off)			
£0		Total Cost (PV)	£	
Other key non-monetised costs by 'main affected groups'				

BENEFITS	ANNUAL BENEFITS		Description and scale of key monetised benefits by 'main affected groups'	
	One-off	Yrs		
	£			
	Average Annual Benefit (excluding one-off)			
£		Total Benefit (PV)	£	
Other key non-monetised benefits by 'main affected groups' Whilst some professional publishers do archive their own material, all the available evidence suggests that the vast majority of individual publishers of freely available online publications within the UK domain do not. Therefore a substantial proportion of the UK's cultural and intellectual output will be lost forever.				

Key Assumptions/Sensitivities/Risks

Price Base Year 0	Time Period Years	Net Benefit Range (NPV) £	NET BENEFIT (NPV Best estimate) £
----------------------	----------------------	-------------------------------------	---------------------------------------------

Summary: Analysis & Evidence (cont.)				
What is the geographic coverage of the policy/option?	UK and Ireland			
On what date will the policy be implemented?	TBC			
Which organisation(s) will enforce the policy?	NA			
What is the total annual cost of enforcement for these organisations?	£ NA			
Does enforcement comply with Hampton principles?	Yes			
Will implementation go beyond minimum EU requirements?	Yes/No			
What is the value of the proposed offsetting measure per year?	£			
What is the value of changes in greenhouse gas emissions?	£			
Will the proposal have a significant impact on competition?	No			
Annual cost (£-£) per organisation (excluding one-off)	Micro	Small	Medium	Large
Are any of these organisations exempt?	No	No	N/A	N/A
Impact on Admin Burdens Baseline (2005 Prices)			(Increase – Decrease)	
Increase of £	Decrease of £	Net Impact £		

Key:

Annual costs and benefits: Constant Prices

(Net) Present Value

Evidence Base (for summary sheets)

- 6.1 This impact assessment is based on information that was provided by LDAP based on their research on the options for the legal deposit of UK online publications, which are available free of charge and without access restrictions. DCMS have only completed the summary sheets.

Options for Deposit: Summary and Analysis

Summary

- 6.2 LDAP agreed on two possible options for deposit, including both voluntary- and regulation-based approaches. In addition, publisher archiving was considered, as an alternative to deposit schemes.

LDL Permissions-based Harvesting and Archiving (Option One)

- 6.3 This option might be considered as the equivalent of a voluntary code for deposit, in that it will not be subject to regulation, but it presumes that legal deposit Libraries will actively harvest ('pull') material, rather than imposing any duty or encouragement upon publishers to deposit. Thus the main burden of activity rests on the libraries to take periodic copies of websites and documents, then ingest them into an archive for preservation. However libraries may only do so after obtaining copyright permission in writing from the publishers or owners of the material; in responding to these requests, online publishers must give permission on behalf of third party rights-holders.

LDL Regulation-based Harvesting and Archiving (Option Two)

- 6.4 This option presumes that libraries, or agents on their behalf, will actively harvest ('pull') material, rather than imposing any duty or encouragement upon publishers to deposit. Thus the main burden of activity rests on the Libraries to take periodic copies of freely available online publications and documents, then ingest them into an archive. However regulation would allow libraries to do so without needing to contact the publishers or owners of the material.
- 6.5 Regulation will also protect both Libraries and online publishers, to a certain degree, from defamation claims and similar liabilities. However to benefit from these protections, all terms of the 2003 Act and regulations must be applied so that access to material harvested must be restricted to legal deposit library premises. Regulation allows the libraries to harvest more effectively, but will not impose on them any specific requirement, other than a 'reasonable' duty to collect.

Archiving Left to the Market (Option Three)

- 6.6 This option proposes there is no central, public sector initiative, and that archiving would be left to the private sector and market forces. Moreover, Libraries would not actively seek to harvest or build a national archive for freely available online publications. Individual publishers might approach a library to archive material on its behalf; this would be subject to separate negotiations and cost-benefit analysis for the specific circumstances.
- 6.7 For each option, the benefits to stakeholders – including publishers, libraries, and members of the public – have been assessed against the overall administrative costs. In addition, other impacts, such as potential risks, were considered. Taken together, these form the evidence base for determining the most appropriate option for deposit.

Question 13: Do you agree with the analysis of these options? Explain why.

Benefits of a Comprehensive Collection

6.8 The 2003 Act recognises the benefit of collecting and preserving a comprehensive collection of UK electronic publications, including those that are online and freely available. A permanent, comprehensive collection, centrally located and easily accessible to such public stakeholders as researchers, educators, and students enhances UK research and scholarship.

6.9 LDAP assessed the degree to which each option realised this benefit:

1. Permissions-based Harvesting (Option One)	Permission is only successfully obtained in 30% of cases (based on UKWAC figures). Therefore out of 4,500 selected, permission would actually be obtained for only 1,350. The archive will therefore represent only a very small proportion (up to approximately 0.5%) of material in scope and perhaps 99.4% may be lost.
2. Regulation-based Harvesting (Option Two)	Regulation ensures a more comprehensive collection, as permissions are not required. It is estimated that up to 79% of material in scope could be captured and archived within available resources.
3. Archiving Left to the Market (Option Three)	Whilst some professional publishers do archive their own material, all the available evidence suggests that the vast majority of individual publishers of freely available online publications within the UK domain do not. Therefore a substantial proportion of the UK's cultural and intellectual output will be lost forever.

Availability to the Public

6.10 The collection must be made available to future generations of researchers, or else the benefit will be lost, and costs will accrue to the public as well as Libraries and publishers.

6.11 Option One not only freely delivers the publications to the public at legal deposit Libraries' premises, but also has the potential to deliver straight to the users' PCs. However, although not subject to the premises restrictions required by the 2003 Act, the permissions obtained within Option One might include restrictions to access imposed by individual publishers. Not only might this lead to additional cost to libraries, but it would also affect to what extent researchers and others had access to a comprehensive collection.

6.12 Option Two ensures the availability of a permanent and comprehensive collection, as the collecting and archiving of publications is subject to legislation. However, it is a collection subject to the terms of the 2003 Act, restricting the provision of access to within the premises of the Legal Deposit Libraries.

6.13 With Option Three, the uncertainty of the access arrangements and cost to researchers mean that wide availability cannot be guaranteed. In addition, Option Three cannot assure the long-term preservation of an archive, especially when publishers go out of business.

6.14 Finally, with Options One and Two, the nation would be likely to reap economic benefits from activity associated with international interest in centrally located, comprehensive national collections. Especially in the case of Option Two, where the collection would be more comprehensive, the archive would be of value to overseas researchers interested in the UK. Therefore it is likely to warrant their investment of time and money in the UK for subsidiary activities relating to its use as a resource, whether they are academic or business activities.

- 6.15 Although Options One and Three do not impose any regulatory impact on publishers and Libraries, only Option Two, despite premises restrictions, can fulfil the goal of a deposit scheme, as defined in the 2003 Act: to deliver a comprehensive collection, both freely available and centrally located, to future generations.

Question 14: Do you agree with the analysis on making content available to the Deposit Libraries? Explain why. What else needs to be taken into consideration?

Financial Implications of Deposit

- 6.16 These benefits must be balanced against administrative and other costs for the major stakeholders, i.e. Libraries and especially publishers.
- 6.17 The harvesting options for deposit imposes most, if not all, the costs on Libraries. Using the cost model (see Libraries' Key Costs Assumptions below), which is a realistic view of what could be achieved within the limitations of available resources, costs can be calculated for each of the Options:

All Legal Deposit Libraries	1. Permissions-based harvesting	2. Regulation-based harvesting	3. Rely on the market
Average total cost per annum	£496,959	£1,132,286	£ Nil
Material gathered after 10 years (benefit)	77 Terabytes (0.5%)	5,263 Terabytes (79%)	0 Terabytes (0%)
Cost per terabyte	£6,476	£215	n/a

- 6.18 Regulation-based harvesting (Option Two) delivers a larger amount of publications for the lowest unit costs.

Question 15: Do you agree with this costing model? Explain why. What else needs to be taken into consideration?

Other Costs and Impacts

- 6.19 In addition to harvesting costs for Libraries, each Option imposes other costs and impacts on stakeholders.
- 6.20 Option One represents a labour-intensive process for both Libraries and publishers to obtain and respond to copyright permission, and this severely constrains the number of websites and volume of material that can be targeted. For Libraries, there is a relatively high staff cost associated with identifying and contacting publishers. For publishers, there are direct administrative and public relations overheads related to responding to requests from libraries, possibly resulting in higher costs to customers.
- 6.21 As permissions-based harvesting represents only a small proportion of the six million or more potential publications, a greater degree of selectivity is required by Libraries. This adds cost and also has an impact upon the benefits of the archive for future research – any selection decision is based upon the current view of what is important in the present and, with the best of knowledge, in the future.

- 6.22 Responding to permissions requests and seeking permissions from third parties also result in high administrative costs for publishers, as well as costs associated with customer and public relations, all of which may translate to increased consumer costs. According to the publishers' survey conducted by LDAP, publishers generally acknowledged that this would be a costly process, especially if there were no procedure in place or if there were a number of requests to process.
- 6.23 For Option Two, regulation-based harvesting, the process can be much more automated (and tools continue to be developed to facilitate the process) than in permissions-based, but the potentially huge volume of material implies significant extra processing and storage costs. Therefore, Libraries must balance the ideal of collecting all available material in a highly automated fashion against their financial and technical limitations.
- 6.24 But regulation will not of itself impose any direct and specific cost to Libraries, nor require any commitment from the public purse. For publishers, costs are marginal, as they do not have to process permissions and do not have to 'push material' to libraries.
- 6.25 With Option Three, those publishers who do archive do not typically make their archives available for access by researchers as a matter of course, or may charge for doing so. If they do make them available for a fee, it is to be expected that their own costs will be offset by revenue.
- 6.26 However, with this Option, cost burdens spread to include other stakeholder groups, namely business and academic researchers; students; educators; and other members of the public. It imposes a direct cost to the public, not just because it falls short of ensuring a comprehensive collection, but also because archives are spread over a range of publishers with diverse access arrangements. In order to access the range of resources, public stakeholders will incur costs in terms of time, travel and subsistence.
- 6.27 Option Two places the least amount of financial burden on publishers, and incurs costs that can be managed within existing budgets for Libraries, while ensuring a permanent historical and cultural collection for future generations.

Question 16: Do you agree with the analysis of the costs and the impacts of each option? Explain why. What else needs to be taken into consideration?

Risks and other impacts

- 6.28 In addition to gathering evidence on benefits and financial implications for each Option, LDAP also identified specific risks and impacts, concerning civil liabilities, illegal content, third party and dynamic content, as well as potential impact on publisher revenue, all of which also influenced its Recommendation.
- 6.29 Without regulation, any material harvested by the libraries is not subject to the terms of the 2003 Act. Therefore, Option One does not benefit from any of the protections in the Act, such as those for Libraries and publishers against defamation liability.
- 6.30 However, even with regulation-based harvesting, as in Option Two, the results of the Copyright Exceptions Consultation may have an impact on preservation, access and use.
- 6.31 Publishers using third-party content are usually required to take it down once licenses have expired. With permissions-based harvesting, Option One, publishers would be expected to obtain permission for this material to be used within an archive. However, these publishers are potentially liable if the libraries have harvested their publications under regulation (Option Two), including this third-party content. A Notice and Take Down policy and procedure would protect both libraries and publishers.

- 6.32 As the harvesting process is partly automated, the libraries risk inadvertently collecting illegal material. Discussions have already taken place with Internet Watch Foundation (IWF) to identify the levels of risk involved and ways in which they might be minimised. IWF has also given advice on the procedures which should be followed if any such material should be found within the archive. Separate discussions have also been held with the Information Commissioner's Office to address any concerns the Commissioner might have in relation to personal data within the archive. Overall the ICO sees no reason to restrict harvesting activity within the context of legal deposit, nor to require additional controls over access than are implied by the 2003 Act, save for the addendum to the Notice and Take down proposals.
- 6.33 On the subject of permissions-based harvesting, one of the publishers surveyed observed, "Permissions are not an issue. Knowing what information to harvest is the issue. ...because our website content changes regularly, such questions would need to be addressed every time a harvest was made." Publishers are concerned over securing ongoing rights for data or images for sites that were made available free of charge but on a time-limited basis. In the face of dynamic content, Option One does not include the degree of flexibility that is required to gather a large number of publications efficiently, and without placing undue financial burden on publishers and libraries alike.
- 6.34 Publishers, in the survey conducted at the beginning of 2008, expressed concern about the possible impact on their revenues of deposit of (and immediate access to) time-sensitive and other types of revenue-generating publications. As with print deposit, deposit of freely available online publications can include provisions for special arrangements to offset any potential problem (see paragraphs 6.43 and 6.44 on embargoes). Moreover, risks can be minimised by the restrictions to access imposed by the 2003 Act.

Question 17: Do you agree with risks identified here? Explain why. Are there other risks that have not been considered? What would their impact be? Are some of these risks actually not really risks? Why?

Rationale for Regulation

- 6.35 Both Options One (permissions-based harvesting) and Three (archiving left to the market) fail to meet the basic requirement of creating a permanent national collection, freely available at designated sites, as envisaged in the 2003 Act.
- 6.36 Only Option Two, regulation-based harvesting and archiving, would meet this requirement for the following reasons:
- **Rate of Deposit:** The large number of freely available online publications and the rate of increase (about 15% per annum) expected pose a challenge to creating a comprehensive collection in the most efficient way possible. Harvesting provides the most efficient and timely solution for deposit of publications in this category: regulations-based harvesting specifically ensures the greatest amount can be collected to build the most comprehensive collection for future generations.
 - **Proportionality:** LDAP wants to ensure that the costs and scale of publications are proportionate to a Recommendation for regulation. Regulation is an appropriate response to the enormous number of publishers and publications in this category. It is the most cost-efficient method of collection for the libraries and imposes no direct financial or administrative burden upon the publishers.
 - **Existing Harvesting Activities and Infrastructure:** Whilst libraries have not yet been able to harvest widely across the whole UK domain, much of the technical infrastructure and expertise that they would need is already in place. They have practical experience of permissions-based harvesting for the UK Web Archiving Consortium. In addition, the International Internet Preservation Consortium (IIPC), of which the BL and NLS are members, is developing tools, infrastructure and the

policies for addressing the challenges of collection and preservation internationally. Moreover, knowledge gained from other national libraries that already carry out domain-level harvesting (e.g. in France, Denmark, New Zealand and Australia) will also assist its implementation in the UK. Indeed, the UK legal deposit libraries estimate that they could begin harvesting on this basis, up to 80% in scale, within existing resources and within a very short time of implementation.

- **Civil Liabilities:** Regulation would protect both publishers and libraries from such civil liabilities as copyright infringement and defamation. However, there are potential risks that may have significant impacts, if not eventual costs.

Question 18: Do you agree with LDAP's recommendation to regulate for this content? If not, what should be done instead?

Policies for Deposit, Access and Use

Restrictions on activities in relation to freely available online publications

6.37 LDAP recommends the following policies regarding provision for access and use as required by Section 7 of the 2003 Act:

- No limitations are imposed on access in the specific areas of time and circumstance of use (7.4 (b)) and number of readers (7.4 (d)). LDAP recommends that the standard terms of access already implied by the 2003 Act are correct and that no changes are necessary. Thus access must be confined to readers (and staff) using terminals, screens or devices that are controlled by the Libraries, and whilst they are on the Libraries' premises.
- LDAP does not seek to extend the conditions of access in the specific areas of time and circumstance of use (7.4 (b)) and number of readers (7.4 (d)).
- Libraries must not be permitted to adapt or change the intellectual content of the material itself. The original 'look and feel' must also be preserved as far as possible (subject to technical limitations).
- In the case of 7.2 (b), the regulation must allow for a single copy harvested by one Library to be made available in all six Libraries and therefore to be regarded as being the equivalent of depositing six copies (i.e. one per LDL) for storage, preservation and access purposes.
- For preservation purposes, legal deposit Libraries must be allowed to transfer publications from one technology platform to another in the event of upgrades to technology or where changes of systems are required.
- Libraries must not be permitted to lend or transfer the material, except in the case of other legal deposit libraries.
- Libraries must be permitted to copy material for the purposes of preserving the material and ensuring that access can continue to be provided, but not in order to increase access.
- Readers must not be permitted to download or make electronic copies of any material.
- Readers must be permitted to make a printed copy [on paper] of the material.
- LDAP requests that there be a note on 'Removal of Access' (or 'notice and take-down') such that, for material in their collections that is subject to a court injunction or judgement, Libraries, subject to notice, will comply with the courts by denying all access to all readers for an appropriate period, which might perhaps be for many years (this does not include deletion or disposal of a web page).
- In the case of 7.4 (f), Libraries must be permitted to dispose of materials prior to incorporation into the collection, defined as the point before which readers have access.
- Moreover, Libraries must not be permitted to dispose of materials once access to readers has been provided.

- 6.38 Regarding subsection 7.4 (b),
- Legal deposit Libraries, while not required to grant embargoes, may consider requests based on demonstrable financial prejudice to the requestor, with reference to the specific conditions of access.
- 6.39 Regarding (7.4 (c)), the Panel would like the Act's definition of 'reader' and 'relevant person' to be included where appropriate. (See 2003 Act Section 5)
- (a) "reader" means a person who, for the purposes of research or study and with the permission of a deposit library, is on library premises controlled by it;
- (b) "relevant person" means –
- (i) a deposit library or person acting on its behalf;
 - (ii) a reader;
- (d) references to a deposit library include references to the Faculty of Advocates."

Policy Background

- 6.40 For discussions on Section 7, LDAP was guided by what is permissible in the 2003 Act as well as in print deposit practice as initial points of reference for policy development. These considerations resulted in policy allowing the reader to print, but not download or make electronic copies ('digital copying'). Libraries will be allowed to make electronic copies for preservation and continued access, but not for lending and transferring. Further deliberation on the subject of printing included whether there should be any limits to the amount or proportion of material that a reader may print out on paper. LDAP acknowledged that this issue might need further work; while some LDAP members felt that no restriction is necessary for this category of material, as it is originally made freely available by the publisher, others felt that a lack of any restriction may have wider ramifications. However, all LDAP members agreed that, in keeping with print deposit practice and in acknowledgement of the nature of the material, a reader should be permitted to make printed copies.
- 6.41 In the matter of disposal (f), while the 2003 Act does not prohibit disposal for print publications, it does contain an express prohibition on disposal for non-print publications. Therefore, LDAP recognises that it is necessary to specify the right to do so in the secondary legislation, for instance of collection of duplicate and partial freely-available online publications. LDAP feels that it is possible to articulate a limited provision of such rights to dispose, perhaps in a guidance note accompanying regulation as it is arguable that duplicate and partial delivery of documents is not effective 'delivery of relevant material' under the 2003 Act and cannot be perceived as contravening the requirement not to dispose, as in the first case one copy has already been collected and will be retained, and in the second instance a representative or complete copy has not been effectively collected and could not therefore be considered as 'disposed'.
- 6.42 LDAP sought to distinguish between harvested publications already incorporated into the permanent collection and those yet to be incorporated. Generally speaking, the restrictions against disposal would cover only those publications part of the permanent collection, and that duplicate or incomplete copies would be disposed of before becoming part of the permanent collection.
- 6.43 With regard to 7.4 (b), LDAP's deliberations on the subject of embargoes originally began with defining and analysing the nature of the publications freely available, free of charge and without access restriction, the collecting or harvesting of these publications, and costs that would be incurred by publishers as a result of such activity. In considering cost, LDAP looked to the BRE guidelines in order to have a better understanding of what costs should be included. Taking into consideration the types of costs highlighted by the BRE and the relative passive role of publishers in the harvesting activity (see description of harvesting above), LDAP's initial assumption was that as no real (admin) cost was incurred by publishers, and because the material was already freely available on the web, there was no

requirement for embargoes. However, in response to subsequent concerns voiced by LDAP members and other stakeholders, LDAP deliberated on two measures that would provide comfort to publisher concerns:

- Embargoes, the well-established practice for printed publications, in which a publisher of material that is, for example, of high commercial value but sold in small numbers, may contact the Libraries and negotiate a temporary embargo upon access to deposited items for a period, typically 1-2 years. These embargoes and their conditions would be reviewed on a case by case basis. They do not prohibit the collection of material nor eventual access to the material. As a result of considering the print precedent, LDAP was persuaded that embargoes were a reasonable policy to apply in the case of the harvesting of freely-available online publications.
- Opt-outs, the practice of allowing publishers of certain types of materials to be exempted from the collection and preservation of their freely available publications, and for this exemption to be considered as part of the policy recommendations made by LDAP. Upon further consideration, LDAP felt that opt-outs were not a reasonable policy to apply in the case of the harvesting of freely-available online publications.

- 6.44 LDAP has considered both options in detail, guided by current practice, research on business models and revenue streams, as well as discussions with publishing stakeholders. While none of these deliberations has produced any causal link between the practice of legal deposit of freely available online publications and the challenge to revenue streams, LDAP felt strongly that publishers' real concerns for future impact on revenue streams should be accommodated. In the specific case of freely available online publications, legal deposit process provides for a much more controlled environment for access than readers' own homes (where this material would be accessible in the first instance). The resulting recommendation is a policy for individual publishers to request embargoes on a case by case basis, as an acknowledgement of these concerns, balanced with the mandate of the collection and preservation of these materials set out by the 2003 Act.
- 6.45 In addition, it is felt that optional embargoes do not compromise the integrity of a regulation (as does allowing certain publication types to opt-out) , nor the "future-proofing" of that regulation (i.e., the in-built flexibility within a regulation to acknowledge the changing nature of publications as a result of technology and business models dictated by it). This concern for future-proofing has been the basis of the LDAP's considerations, beginning with the EPS consultation.
- 6.46 With regard to material currently held by the Libraries which has been deemed potentially prejudicial and therefore inappropriate for public access by the courts, the libraries would presently not be notified of any resulting injunction or judicial direction against the publisher. LDAP understands that the Libraries are aware of risks and current advice on this issue. With regard to inadvertent harvesting of online publications, the publication and/or dissemination of which may be an offence (e.g. under the Terrorism Act 2006 and the Sexual Offences Act 2003), a case may be made to grant libraries an exemption from criminal liability (pursuant to the power to do so under the 2003 Act), for both the act of harvesting and retention of the publication if the Panel and its library members see such immunity as a desirable objective.

**Question 19: Do you agree with LDAP's proposed method for depositing of content?
If not how else could this be done?**

**Question 20: Do you agree with LDAP's analysis of access provisions? Explain why.
What other options are there?**

Libraries' key cost assumptions

- 6.47 The category definition and territoriality rules govern what may be collected. Within these, this model assumes that the UK web space is defined as all .uk domains registered by Nominet (6.1 million in mid-2007) plus approximately 50,000 other domains which can be readily identified as published in the UK. The numbers will continue to grow by 17% per annum until 2011, then by 15% till 2016¹⁰.
- 6.48 However 35% of domains are inactive¹¹, i.e. registered but not live, or where static content can be 'de-duplicated' after a first snapshot. A further 2.5%¹² are primarily 'deep web' or protected sites outside the scope of this category. Overall, the number of websites in scope is therefore estimated as 3.9 million in 2007 rising to 14.6 million in 2016.
- 6.49 The average size of each site has been growing significantly each year. However this model assumes that most audio-visual content, one of the major causes of growth, is out of scope and therefore a more modest 5% growth per annum is appropriate. The average size also varies dramatically, from circa 5 megabytes for 80% of sites to 1 gigabyte for 0.5% of sites; this model assumes a weighted average of 25 megabytes. However, in permissions-based harvesting selection tends to favour the larger websites, so this model assumes an average 180 megabytes in that option.¹³
- 6.50 The cost of storage includes built-in redundancy to ensure safe preservation of the archive; content is replicated (with security measures) across three nodes in the BL/NLW shared infrastructure and across 2 nodes in the NLS infrastructure. However the real cost of storage per terabyte has fallen by more than 30% per annum over the last 20 years and is expected to continue falling by 25% per annum until 2016.¹⁴
- 6.51 The two infrastructures have already being designed and built – apart from certain elements – by the libraries and will be used to store other digital or digitised collections as well as legal deposit material. Therefore this model focuses only on the incremental systems costs – including renewing equipment every three years – salaries, pensions, NI, other staff related costs, plus allocations for general support and overheads, that are required to collect and preserve 'free web' material.

Question 21: Do you agree with these cost assumptions? Explain why. What needs to be included or excluded?

¹⁰ Assumption based upon historic UK growth data from Nominet, historic .com/.org/.net data from www.zooknic.com and <http://www.oecd.org/dataoecd/56/34/32996948.pdf>

¹¹ Assumption based upon de-duplication experiences reported by Royal Library of Copenhagen

¹² Source = <http://www.press.umich.edu/jep/07-01/bergman.html>

¹³ Assumptions and estimates based upon historic data from UKWAC and national libraries in Denmark, Australia and New Zealand

¹⁴ Conservative assumptions based upon Moore's Law, Kryder's Law and Kurzweil quoted in various articles, plus sources: <http://www.sciam.com/article.cfm?articleID=000B0C22-0805-12D8-BDFD83414B7F0000&ref=sciam&chanID=sa006> and http://www.itrs.net/Links/2006Update/FinalToPost/01_SysDrivers_2006UPDATE.pdf

Specific Impact Tests: Checklist

Use the table below to demonstrate how broadly you have considered the potential impacts of your policy options.

Ensure that the results of any tests that impact on the cost-benefit analysis are contained within the main evidence base; other results may be annexed.

Type of testing undertaken	Results in Evidence Base?	Results annexed?
Competition Assessment	Yes/No	Yes/No
Small Firms Impact Test	Yes/No	Yes/No
Legal Aid	Yes/No	Yes/No
Sustainable Development	Yes/No	Yes/No
Carbon Assessment	Yes/No	Yes/No
Other Environment	Yes/No	Yes/No
Health Impact Assessment	Yes/No	Yes/No
Race Equality	Yes/No	Yes/No
Disability Equality	Yes/No	Yes/No
Gender Equality	Yes/No	Yes/No
Human Rights	Yes/No	Yes/No
Rural Proofing	Yes/No	Yes/No

Annex E: Further Details on Territoriality

Further Information on 'Published in the UK' (Territoriality) for Online Works

- 7.1 Sections 5.3 to 5.6 of this recommendation propose a way of defining which publishers may be affected by the regulation for this category of material. This additional information about the proposed definition illustrates how it may be interpreted in practice.
- 7.2 A domain name is an identification label that defines a realm of administrative autonomy, authority, or control in the internet. A single website may consist of resources with just one domain name such as 'anyorganisation.co.uk' or might also include resources with other domain names, such as one website linking together 'anyorganisation.co.uk', 'anyorganisation.com' and 'anyorganisation.biz'.
- 7.3 An online 'publication' may include any electronic file, document, image, data, web page or web application. It also includes audiovisual content (sound and video) where this represents a part, but not the whole, of the complete work; for example a news website with sound or video clips is included, but cinema films, TV and radio programmes, music albums and songs, even if provided via an online service such as BBC iPlayer, are not.
- 7.4 'Publisher' includes individuals, companies, groups and organisations, and is the person or body which makes a publication available, either directly or through an agent acting under its authority. This excludes persons or bodies such as internet service providers whose sole involvement is in providing the network infrastructure or the means of publishing, but who have no direct responsibility for the content.
- 7.5 'Making available to the public' excludes private intranets.
- 7.6 'UK electronic address' refers to any domain name that is associated geographically with the United Kingdom, or with part of the United Kingdom. Therefore all publications with **.uk** domain names are included. Any publications with the proposed **.sco** and **.cym**, domain names or any other domain names which might be created in future for UK regions, cities or towns will also be included. Because these have a UK electronic address, it is not necessary to identify where the publisher is physically domiciled.
- 7.7 Some UK publishers prefer to use **.com**, **.biz**, **.tv** or other generic or international domain names. If the publisher has a UK 'physical address' they are still included, irrespective of the domain name being used. This encompasses publications made available by a publisher who is either domiciled in the UK, or a multi-national with an address in the UK, or the UK branch or subsidiary of a non-UK organisation.
- 7.8 However publications are excluded if they are made available by a publisher outside the UK and using a non-UK electronic address, even though the content of such publications might be about the UK or relevant to the UK, and even though the publication might be viewed or purchased by someone in the UK. For example an online Travel Guide to London being used by a person in the UK is not in scope if it is published by a US organisation via a **.com** website.
- 7.9 Internet Protocol (IP) addresses may indicate the location of a computer or server connected to the internet and can therefore be a useful way of identifying whether a publication might potentially be within scope. However the IP address on its own is not sufficient; the publication must either be within a UK domain as above or the publisher must have a physical address within the UK.

Annex F: Further Details on Harvesting Process

Explanation of the Harvesting Process

- 8.1 The Panel's goal has been to propose a deposit process which is cost-efficient for the legal deposit Libraries and which also imposes no administrative cost burden upon publishers. Because it potentially involves such a large number of publications, the Panel has recommended that libraries pull them directly from the Web using an automated process in which no action is required of publishers. This process will use a software tool ('harvester') to crawl relevant web domains.
- 8.2 An initial seed list of Uniform Resource Locators (URLs) will be loaded into the harvester by library staff. These will usually be URLs for the home or root pages of web domains that are within the scope as recommended by the Panel.
- 8.3 For each URL, the harvester will issue an electronic request to the publisher's web hosting server for delivery of a copy of the page or file. Each request will include information which identifies:
- the Internet Protocol (IP) address of the harvester issuing the request
 - the URL for the page or file requested
 - a 'user-agent string' which identifies the library controlling the harvester and the fact that it is a harvesting request
 - the URL for a web page containing details of how to contact the library, plus contextual information about legal deposit and the terms of the regulation.
- 8.4 The web hosting server responds automatically, delivering a copy of the page or file to the harvester. Once the copy has been delivered to the harvester, it may then be incorporated into the library's archive collection.
- 8.5 Essentially the same process, albeit with different information contained in the 'user-agent string', underlies all browsing activity by every web user; web publishers will not need to make any systems changes or undertake any action to facilitate this.
- 8.6 Website owners may not choose to log this information, but the general practice is certainly to log the user-agent string; many use this information actively to tailor content accordingly for different users, e.g. for mobile phones as opposed to computer browsers.
- 8.7 Libraries will set rules and parameters for the harvester to ensure that there is no harmful impact upon the performance of the web hosting server:
- Only web pages and documents that are publicly and freely available will be requested; harvesting will not go anywhere that is not public.
 - Web pages and documents will only be harvested periodically; the Panel's cost estimates were based upon an assumed average of twice a year.
 - When multiple requests for different pages and files are issued to the same web hosting server, a generous interval between each request will safeguard against any risk of using up bandwidth or overloading the server.

- The harvester will not obtain any content that is protected by a firewall or by any kind of barrier such as username/password protection.
- The harvester will not request any pages or documents that do not have web links to them; therefore any pages or files which are not part of the public website cannot be requested.

8.8 The harvester will automatically follow links from the home or root page to the next levels down within the same domain, issuing a separate request for each page or file.

Annex G: Questionnaire

10.1 The following questions relate to the proposals set out in Annexes A to F.

Question Number	Question	Page
1	What are your views on the options considered for this content?	10
2	Are there any other options that should have been considered? If so what are they?	10
3	Do you agree with the analysis of these options? Explain why.	10
4	Do you agree that harvesting provides the most efficient and timely solution for deposit of publications in this category? Explain why.	10
5	Do you agree regulation is the most cost-efficient method of collection for the Libraries and imposes no direct financial or administrative burden upon the publishers? Explain why.	10
6	Do you agree that this is an appropriate definition for the type of publications that should be included in scope for regulations? Explain why. Is there anything else that should be included in this definition? Is there anything that should be excluded from this definition?	12
7	Do you agree with the territorial definition of the UK web? Explain why. Is there anything else that should be included in this definition? Is there anything that should be excluded from this definition?	13
8	Do you agree with this analysis of the UK Web Domain? Explain why. What do you think the impact of your analysis would be?	13
9	How do you see a Deposit Library driven system of web harvesting interfacing with a publisher driven duty to deposit under the 2003 Act?	14
10	How could Deposit Libraries most efficaciously ensure a comprehensive body of eligible content is deposited?	14
11	Do you agree with this costing model? Explain why. Are there costs that need to be factored in or excluded?	15
12	Do these assumptions adequately reflect the financial burden of publishers? Is there anything that needs to be included or excluded?	16
13	Do you agree with the analysis of these options? Explain why.	25

Question Number	Question	Page
14	Do you agree with the analysis on making content available to the Deposit Libraries? Explain why. What else needs to be taken into consideration?	27
15	Do you agree with this costing model? Explain why. What else needs to be taken into consideration?	27
16	Do you agree with the analysis of the costs and the impacts of each option? Explain why. What else needs to be taken into consideration?	28
17	Do you agree with risks identified here? Explain why. Are there other risks that have not been considered? What would their impact be? Are some of these risks actually not really risks? Why?	29
18	Do you agree with LDAP's recommendation to regulate for this content? If not, what should be done instead?	30
19	Do you agree with LDAP's proposed method for depositing of content? If not how else could this be done?	32
20	Do you agree with LDAP's analysis of access provisions? Explain why. What other options are there?	32
21	Do you agree with these cost assumptions? Explain why. What needs to be included or excluded?	33

We can also provide documents
to meet the specific requirements
of people with disabilities.
Please call 020 7211 6200 or
email enquiries@culture.gov.uk



department for
culture, media
and sport

2-4 Cockspur Street
London SW1Y 5DH
www.culture.gov.uk