



LORD CHANCELLOR'S
ADVISORY COUNCIL ON NATIONAL RECORDS AND ARCHIVES

Proposal on the Collection and Preservation of UK Offline and Microform Publications and UK Online and Microform Publications and UK Online Publications Available Free of Charge and Without Access Restrictions

1. Background on the Advisory Council on National Records and Archives

1.1 The function of the Lord Chancellor's Advisory Council on National Records and Archives is to advise the Lord Chancellor, and through him other ministers as appropriate, on all matters relating to records and archives in the United Kingdom and in particular in England and Wales. It is chaired by the Master of the Rolls.

1.2 The Advisory Council on National Records and Archives embraces two further bodies: the Advisory Council on Public Records, established by s1(2) of the Public Records Act 1958(2), and the Advisory Council on Historical Manuscripts, established by the Lord Chancellor to advise him on matters formerly the responsibility of the Royal Commission on Historical Manuscripts. All three bodies share the same members.

1.3 The terms of reference of the Advisory Council on National Records and Archives are to consider and advise the Lord Chancellor upon:

- Major objectives, programmes and policy changes for The National Archives
- Proposed legislation affecting The National Archives
- The implications for records and public services of any proposed change to the status of The National Archives
- Any subjects brought to its attention by the Lord Chancellor

2. The Advisory Council's overall views on the consultation

2.1 The Advisory Council welcomes the opportunity to respond to this consultation. It has a particular interest in United Kingdom online publications available free of charge and without access restrictions. The Internet has radically shifted the boundaries between published and unpublished material. The World Wide Web is being increasingly used to disseminate to the public information which would previously have never been cost-effective to publish in printed form. It thus subsumes material which would formerly have been of an archival (unpublished) nature. The Advisory Council therefore supports Government's proposals to pursue a legislative option for the archiving of online publications available free of charge and without access restrictions by means of Statutory Instrument pursuant to the Legal Deposit Libraries Act 2003.

2.2 For further information about our response, please contact Dr Grazia Zaffuto, Advisory Council Secretary on Tel: 0208 392 5377 or by email:

advisorycouncilsecretary@nationalarchives.gsi.gov.uk

The Advisory Council's response to the consultation questions

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

3.1 The Advisory Council believes that Option 2, Legal Deposit Libraries Regulation-based Harvesting and Archiving offers the most feasible approach for ensuring that the nation's published output falling within the category covered by this consultation is collected systematically and as comprehensively as possible. Moreover, Option 2 has scope to support and interface with other UK-based harvesting and archiving programmes.

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

4.1 Rather than suggest further options, it is agreed that the legal deposit libraries should continue to work closely with other key stakeholders and partners with an active interest in web archiving, so as to share expertise and avoid duplication of effort. These would include other memory institutions such as The National Archives.

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

5.1 The analysis is sufficiently robust to enable the Advisory Council to recommend Option 2 as the best approach for preserving that proportion of the UK's web content which is available free of charge and without access restrictions, and within the specified funding constraints of the legal deposit libraries. The Advisory Council considers that Government has demonstrated that Option 2 is the most cost-effective method of archiving the greatest number of websites, whilst also minimizing the burden upon publishers and addressing civil liability issues.

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

6.1 The Advisory Council entirely agrees that harvesting provides the most efficient and timely solution for deposit of publications in this category. The Council agrees with LDAP's analysis that, in the light of the direct experience of the legal deposit libraries thus far, permissions based harvesting (Option 1) would be an ineffective solution on account of difficulties with obtaining permissions and because of the higher cost per unit for publishers and Libraries. The Council agrees that archiving left to the market (Option 3) would not be an effective solution, as indicated by the available evidence.

7. 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.

7.1 The Advisory Council agrees that regulation is the most cost-efficient method of collection for the Libraries and imposes no direct financial or administrative burden upon the publishers. Self-evidently this is the case for publishers, since there is no loss of revenue and no administrative burden of active deposit; and the Libraries themselves have estimated they could begin harvesting up to 80% in scale within existing resources and within a very short time of implementation.

8. 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?

8.1 The Advisory Council agrees that, within the constraints imposed by Section 1(5) of the Legal Deposit Libraries Act 2003, the definition is appropriate for the type of publications that should be included in scope for regulation. The definition encompasses online publications which are available free of charge and accessible without restriction and which can be collected without any requirement for action by publishers, thus avoiding the majority of administrative complications in their collection.

9. 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?

9.1 The Advisory Council agrees with the territorial definition of the UK web because it sets out how the definition may be interpreted in practice and what it includes and excludes. However, although inferred in Annex F, it is not made explicit that this definition has been tested to ensure the automatic capture of content by web crawlers within these parameters.

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

10.1 The Advisory Council is unable to offer any supplementary or alternative quantitative estimates to those which appear in Annex D. However, it does agree that there may be some grounds for thinking them to be a little on the low side (see our response to Question 15).

11. 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?

11.1 The Advisory Council is of the view that, in order to satisfy the interests of both publishers and legal deposit libraries, the regulation should state clearly that web harvesting does not otherwise absolve publishers of their duty to deposit under the 2003 Act.

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

12.1 The Advisory Council supports a robust system that is at once able to capture content at regular intervals, such as the one mentioned in Annex F, paragraph 8.7; and also to take a pragmatic and selective approach that enables Deposit Libraries to prioritise and target particular websites that are continuously updating their content, so that valuable information is not missed.

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

13.1 While the Advisory Council concurs that the cost of “raw” storage has been declining at 30% per annum and is expected to continue at 25% per annum, the real question is the decline in system costs, the tape libraries and arrays in which the storage resides, software and associated technology solutions. These have not been declining at such a rate. It would have been useful for the legal deposit libraries to have provided their cost ranges and more detailed cost assumptions for hardware and software development and replacement.

13.2 With such high volumes of storage, the Advisory Council considers that data communication costs could potentially become an issue. It is unclear whether these have been included in the cost assumptions. More information about the high-speed networking capability of the legal deposit libraries, both individually and collectively, would have been useful in this context.

13.3 We note (Paragraph 5.13) that “They [the costs] 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.” We wonder whether there has been any risk assessment by the libraries that the resources have been undersized and therefore undercosted?

13.4 Resource discovery (Paragraph 5.16) we understand to mean locating and accessing publications within the archive. If this is to be delivered by an existing infrastructure, this should be clarified, and there should be an allowance for additional costs to support the additional data.

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

14.1 In general, we consider that these assumptions do adequately reflect the financial burden on publishers, taking into account the results of the LDAP survey which is referred to, and noting that Government intends to address civil liability issues and that no loss of revenue is involved (since the publications concerned are already being made available free of charge and without access restrictions).

14.2 However, there may be a potential short-term impact on the publishers in some cases whilst the actual harvesting is in process. The demand on the websites may impact the publishers' own expected service levels. This will depend on the volume of data being harvested. More information on this point would have been helpful.

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

15.1 The Advisory Council agrees with the analysis of the options. See its response to Questions 1 and 2 above.

16. Question 14: Do you agree with the analysis on making content available to the Deposit Libraries? Explain why.

16.1 The Advisory Council agrees that the proposed regulation is necessary in order to preserve and make accessible this significant sub-set of the nation's published heritage. Additionally, the Council is also of the view that web material should be made available in its original format, as per those web archive collections at the British Library for instance, which are already accessed via the web itself.

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

17.1 The cost aspects have been covered above, in our response to Question 11.

17.2 Using figures in paragraphs 5.8 and 5.9, it appears there are three million domain names, as of 2007. This means that each domain has 1.3 publications, on average. This appears to be a rather small ratio.

17.3 The Advisory Council is unsure if the growth of new publications of 15% per annum is a high enough figure.

17.4 While the definitions are clear as to what will be included, translating this into volumes may underestimate the developing multi-media world over the next decade.

17.5 Costs and definitions need to be aligned, especially when this includes high growth publication types.

18. 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?

18.1 Option 2 becomes more deterministic and offers the fewest barriers to collecting the publications and building the historical record of the country. Our comments on costs are set out in our response to Questions 11 and 12.

19. 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?

19.1 The Advisory Council agrees with the risks identified, although it has some difficulty in understanding publishers' concerns about impact on their revenues of deposit of (and immediate access to) time-sensitive and other types of revenue-generating publications, as this consultation is concerned solely with publications that are made available by publishers free of charge and without access restrictions.

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

20.1 The Advisory Council agrees with LDAP's recommendation to regulate for this content. It considers that there is no viable alternative means of creating a core national web archive.

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

21.1 The Advisory Council agrees with LDAP's proposed method for depositing content but it also recognises that other specialist web collections are made available to researchers and members of the public by public sector and other agencies besides the legal deposit libraries. For this reason, it is important that these libraries maintain good relations with these other memory institutions to ensure an optimal approach to the capture and preservation of the nation's digital heritage.

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

22.1 The Advisory Council notes that access to the material will not be provided beyond the physical confines of the legal deposit libraries. We appreciate that this is in line with the provisions of Section 7 of the Legal Deposit Libraries Act 2003. However, researchers and members of the public will doubtless find it ironic and frustrating that content which, prior to archiving, was freely available to anybody throughout the world will now become so restricted.

22.2 The Advisory Council further believes there is a misalignment of policy and practice where it concerns the proposal to allow readers to make printed copies of website material but not

download or make electronic copies of any material for non-commercial purposes. Such an approach is restrictive and contrary to the needs of researchers.

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

23.1 The Advisory Council agrees with the cost assumptions subject to the reservations noted above.

**Advisory Council on National Records and Archives
1 March 2010**