
Department of Culture Media & Sport - Proposal on the Collection and Preservation of UK Offline and Microform Publications and UK Online Publications**Response by the Wellcome Trust**

January 2010

1. The Wellcome Trust is the largest charity in the UK. It funds innovative biomedical research, in the UK and internationally, spending over £600 million each year to support the brightest scientists with the best ideas. The Wellcome Trust supports public debate about biomedical research and its impact on health and wellbeing.
2. Part of the Wellcome Trust, the Wellcome Library is one of the world's major resources for the study of medical history and we also provide access to a growing collection of contemporary biomedical information resources relating to consumer health, popular science, biomedical ethics and the public understanding of science.
3. The Library, along with the JISC and the British Library, was a founder of the UK Web Archiving Consortium (UKWAC). The Library has contributed online publications to the UK Web Archive since 2004, archiving material on a permission-secured basis.
4. The Wellcome Library supports the Legal Deposit Advisory Panel's (LDAP) recommendation of regulation based collection of UK online publications. We have some reservations about access to archived publications and to the restrictive aspects of some of the proposed regulation.

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

5. The Wellcome Library supports the Legal Deposit Advisory Panel's LDAP recommendation Option 2 (Legal Deposit Libraries Regulation-based Harvesting and Archiving). The Library also supports the concept that 'agents' may act on behalf of Legal Deposit Libraries.

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

6. The Library believes that regulation is a practicable and workable approach to gathering UK online publications for preservation. Legal Deposit Libraries (LDLs) are the bodies with responsibility to secure published material of national cultural importance. LDLs also have relevant expertise and experience in identifying material for inclusion in national collections, expertise in its long term management and the infrastructure for its dissemination. They are a trusted source of information and knowledge. It is appropriate that this responsibility is extended to UK online publications.
7. Harvesting and archiving UK online publications is a technically challenging activity. Within the UK there is significant expertise in LDLs both in the form of subject expertise and in the practice of archiving online publications.
8. That said, there will be occasions when assistance is required – either for technical reasons, or for reasons of domain expertise – and thus it is important that the LDL can make use of "agents" to support the gathering of UK online publications for preservation.

Question 6, 7, 8 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?

9. The Wellcome Library believes that the definition of UK Territoriality and UK Domain is too restrictive. We are generally comfortable with the list of publications to which the proposed regulation applies, but not with the way in which those documents can be archived. The proposal does not account for the vagaries of gathering UK online publications. Web archiving is an imprecise art, dependent upon technology and only able to respond reactively to changes in Internet technology. It is also undertaken 'blind', in that little knowledge can be gained about online publications and their technical properties before they are gathered. The effect of the proposed regulation would be to make the process of archiving material more challenging than necessary. This would drive up cost and hinder the efficiency of the process.
10. For technical reasons it is not always possible to reliably restrict gathering of online publications to that of a single domain. The many ways in which HTML code can be used, and websites constructed, means that in some cases near-line content is inadvertently archived, e.g. book jacket images from Amazon in the US, documents from non-UK government websites, or material from sites that provide a service to the single domain being archived, e.g. content from Google Analytics. This is known as 'collateral archiving'.
11. The effort required to remove, or exclude this 'collateral archiving' content may be disproportionate to the quantity of material inadvertently gathered. Any regulation should allow such collateral content to be retained where reasonable steps have been taken to gather only material from a selected domain. To be clear, if the software used to gather online publications is set only to gather www.mysite.co.uk – and somewhere on that site, there is a document from a non UK site, then that document should be retained in the archive. To go through every single page archived – to check that no collateral content had been harvested – would be too expensive, and disproportionate.

Question 20 Policies for Deposit, Access and Use.

12. The Library does **not** agree with LDAP's recommendation that access (to archived online publications) must be confined to readers (and staff) using terminals, screens or devices that are controlled by the libraries and whilst they (readers) are on the Library premises. (6.37) The Library believes that in this regard LDAP is out of step with ways in which online publications are used.
13. The recommendation for access is also at odds with the Government's 'Digital Britain' final report¹. The report's vision for ensuring that the UK is at the leading edge of the global digital economy and its desire to maximize the social and economic benefits from digital technologies is severely undermined by restricting access to a permanent record of the very content of the digital environment.

¹ See http://www.culture.gov.uk/what_we_do/broadcasting/6216.aspx

14. The Wellcome Library believes that UK online publications gathered under legal deposit regulation by a Legal Deposit Library should be made available in the same way that they were originally. That is, free of charge and without access restrictions via the internet to any user in any location.
15. Arguments in favour of this are articulated below:
- Storage and preservation of online publications is technically complex, time consuming and carries a high cost. Without the widest possible public dissemination of this material the meagre access afforded to a few individuals in library reading rooms will render the maintenance of archived material uneconomic. Only wide-scale use of the material can justify the effort and cost involved and justify its status and national cultural heritage.
 - Social, scholarly and scientific communication is increasingly conducted digitally. In a global knowledge economy there is no advantage in placing artificial access restrictions on information or knowledge. Online publications were available free of charge and without access restrictions at the time they were published. They should be freely available and without access restrictions when stored in an archive.
 - Digital data offers unique ways in which it can be used for research purposes, ways that are not possible with 'analogue' publications. Multiple data can be concurrently searched, be fully indexed or text mined by automated processes. Thus large amounts of data can be quickly and efficiently processed in the creation of new knowledge.
 - If archived sites are effectively placed in a 'dark archive' – with access only being made available to those who physically visit a LDL – there is a risk that errors in harvesting (e.g. not all pages/images etc were successfully archived) will not be identified. This, in time, will impact of the usefulness of the archive for research purposes.