



# Legal deposit and collection development in a digital world

Nicholas Joint  
*Editor, Library Review*

## Abstract

**Purpose** – To compare and contrast national collection management principles for hard copy deposit collections and for digital deposit collections.

**Design/methodology/approach** – A selective overview and summary of work to date on digital legal deposit and digital preservation.

**Findings** – That the comprehensive nature of traditional print deposit collection often absolves national libraries from the more intractable problems of stock selection; whereas the difficulty of collecting the entire national digital web space means that intelligent selection is vital for the building of meaningful digital deposit collections.

**Research limitations/implications** – These are indicative and partial insights based on small scale interrogation of trial digital deposit collections: the issue of collection development and selection biases in digital collection building needs greater in-depth research before hard and fast recommendations about collection management criteria can be arrived at.

**Practical implications** – The principles outlined may offer practitioners in national libraries some useful insights into how to manage their digital deposit collections.

**Originality/value** – This paper emphasises the social and political aspects of digital deposit issues, rather than the legal or technical aspects.

**Keywords** Digital libraries, National libraries, Collections management

**Paper type** Viewpoint

## Introduction

During a previous part of my career, when I was a cataloguer at the British Library, I occupied a cavernous office. It had to be large – I shared the space with part of the backlog of uncatalogued material given to the British Library under the legal deposit system. I enjoyed my time at the British Library, but I can still recall the slow encroachment into my office area of the legal deposit backlog. After some years, I left for another post, thus abandoning my shrinking patch of workspace – quite possibly to be engulfed entirely by this expanding cataloguing backlog. *Après moi le deluge*, indeed! In retrospect, this relinquishing of workspace may have been my most important personal contribution to the task of containing our national bibliographic output.

However, looking back, in some ways we had it easy in the national library. Because, above all, we did not have to make any decisions about stock selection: if something was published, we received it automatically, and then we would try and catalogue it. The size of the task was immense, but there was little doubt about what we had to collect.

The application of legal deposit principles to the digital world is, however, not so simple. Since the outline creation of a statutory framework for legal deposit of non-print materials with the Legal Deposit Libraries Act (2003) in the UK, we in the British library and information science (LIS) profession, like our colleagues elsewhere, have had to wrestle with the concept of digital legal deposit. Now, some 3 years down the line, we have some sort of idea about how to address this challenge.



### Traditional legal deposit models

One of the most straightforward approaches to digital legal deposit is to do it as we have already done it. After all, most electronic information products are publications in parallel – so national libraries merely have to ask governments to mandate electronic publishers to give them files of the e-versions of those printed outputs which are already subject to traditional legal deposit, and then we can grapple with the associated national deposit problems of long-term storage and digital preservation. The e-journal deposit system at the Koninklijke Bibliotheek, the National Library of the Netherlands, is an outstanding example of this approach (Steenbakkers, 2004), although it is interesting to note that this has been achieved without a national legal deposit act along British lines.

Back in the UK, the government has continued its support for digital collection building by archiving government digital materials, in order to save and preserve our electronic national heritage. Within this sphere, the UK National Archives have been exemplary in their approach to preserving “born digital” national information artefacts. Where government is creating outputs in purely electronic form, then the digital artefact alone must be stored and preserved. The compelling framework for building these national collections is not one of legal deposit, but rather the idea that we must maintain a public record of the workings of government. There is clearly therefore an overlap with the philosophy of legal deposit, in that both draw on a sense of national mission: the National Archives must keep the national record alive as a public good and public duty.

From a collection management point of view, however, the role of the legal deposit library is the more straightforward. The National Archives have it a bit tougher: realistically only 5 per cent of the output of government can be preserved and an expert approach to selection for preservation is required (National Archives Acquisition Policy, 1998).

So as we move from parallel print/digital publication to the enormous government output of born digital items, the task of selection gets more complicated. However, the biggest challenge lies in building a digital archive of informally published born digital material: the unique offerings of the national web space itself.

### Archiving the web

Initiatives such as the UK Web Archiving Consortium (UKWAC, 2006) have created some sort of model for dealing with this, perhaps the ultimate digital heritage challenge. The legislative groundwork is in many ways the least of the problem: much of the web is informally published by those who would be delighted to have their offerings recorded in a national digital archive and the idea of having to compel them by statute to offer their material to a national online collection is redundant. They would jump at the chance. We probably have a greater need for an inverse bill of legal non-deposit for much of the web: this would oblige the authors of meretricious web sites to keep their offering to themselves and share it with no other users of the internet under any circumstance!

That, sadly, would be legalised censorship and quite objectionable – however tempting. So the mixed blessing that is the current creative anarchy of the web must of course remain untouched. But let it be assumed that the regulations on non-print deposit, once formally derived from the 2003 bill, will give full legal entitlement to archive any part of the UK national web collection. How do we then distinguish what

is worthwhile from what is not? This really is the greatest challenge of digital legal deposit.

By comparison, the old dilemma of getting on top of the mountain of national print output is not that hard. Until technically we reach a point where we can, without discrimination, store the national web space in its entirety (which in itself may not give the most coherent result to posterity), digital legal deposit confronts us not with a mountain but a whole mountain range: first, we have to select one or two peaks to keep and preserve for posterity before we even start to grapple with the task of cataloguing and preserving this material adequately. Creating a selection policy for collection development is at present a real headache for digital legal deposit librarianship.

### **The story so far**

To date, much of the effort directed towards solving the problems of digital legal deposit has focused on:

- the output of government,
- “dual nature” or hybrid print/digital items,
- preservation issues (e.g. the Digital Preservation Coalition, 2001).

The end result is a set of digital deposit collections with certain very marked characteristics. These are as follows:

- There is a tendency to preserve the voice of officialdom and the status quo, as in the work of the UK National Archives.
- There is also a tendency to preserve the same dual format print/electronic publications across different countries.
- There is a tendency to explore the preservation problems of dual format print/electronic publications whose print preservation problems are nevertheless well known and solvable, thus expending effort on material that we were always less likely to lose – ideally, this is not where such effort should be directed.

To expand further: if one investigates the holdings of UKWAC (2006) on topics such as the Hutton report or the 7 July 2005 London bombings, the voice of the national digital collection is overwhelmingly the voice of government. Because of that, I would suggest that this national digital collection is not representative of what was the national vision of these defining events in the UK’s history. The alternative perspective to government is difficult to detect – for that you needed to preserve the national web space, with its uncontrolled mish-mash of polemic, politics, truth and trash.

Second, although it is wonderful that a range of national institutions are preserving internationally significant e-journals, in so doing, each national library is preserving largely the same internationally significant e-journal collection. Because digital legal deposit normally only gives a right to offer access within the physical area of the holding collection, this mass international replication of similar collections in different places is acceptable – digital rights regulations should never allow commercially published legal deposit collections to be accessed across a whole national network, because rights holders would either never deposit, or they would be bankrupted.

So it is good to try to ensure that scholars of the future in every country will be able to physically visit their national deposit collections and read the same e-journals in

each reading room of the world. But this digital deposit model is not the same as one that preserves a culture's *unique* national digital output, which will tend to be the anarchic, creative explosion of the national web space in all its mix of glory and goriness.

Third, there are many good reasons for digital preservation projects such as the LIFE project (LIFE, 2006) to focus on dual format print/electronic publications. However, I would suggest that the main value of such life-cycle costing is the “generalisability” of the findings to web-only publications. Most of the e-journals studied will survive quite nicely in their print versions, so the essential outcome will be to apply these findings to the unique national web output which has no print equivalent. Knowing what is the important part of the nation's web-only information landscape that should be preserved is a very pressing question. It is one fraught with legal and political difficulty, because much of the most interesting material to be collected and preserved on deposit may embarrass government, and if untrue, may be libellous or even a genuine risk to national security. For these reasons, working in a legal deposit library has never before been so complex!

### Practical examples

#### *The Hutton report, 2004*

To illustrate these points, it is useful to search UKWAC (2006) for information about the topics mentioned above. At the time of writing, a search for data on the Hutton report produces a multitude of archived hits relating to the official Hutton report website (2004). A few dissenting voices can be found, but they are difficult to trace (for example, the Hutton items from the UK Polling Report, 2004, as preserved on UKWAC). The dissenting web pages which I managed to trace were rather fragmented and hard to understand and when to make more sense of them, I searched for them on their parent web site outside of UKWAC which is still active (UK Polling Report, 2006), the pages had disappeared from the original site. And some other Hutton hits on UKWAC from other sources were simply incomprehensible (see the curious hit on the Dictionary of Slang, 1996–2006, which in some way is linked in UKWAC to the keywords “Hutton report”).

In contrast, the official Hutton report website, which is so well represented on UKWAC, is still very much in evidence elsewhere in its original place on the web, so that the need to archive so much of it on UKWAC seems questionable at this time. And again, this government web site is a parallel print/electronic artefact, so even if none of this official web site were available on UKWAC, the print versions of the Hutton report would be much easier to preserve in the first place, given that these print archival challenges are familiar to us. We might as well dump the Hutton items on UKWAC in favour of relying on the print copies of the Hutton report (Lord Hutton, 2004). But we would also have to accept that the alternative, “web-only” view of Hutton was never preserved. A vital web resource has been lost a mere 2 years since that great controversy, with its tragic human dimension, so dominated the British public imagination.

#### *Event-based web archiving*

By contrast, UKWAC's interesting policy of “event-based” archiving gives a multitude of fascinating views of the later, but not unrelated post-Hutton event, the 7 July London bombings. Their page of archived links entitled “Terrorist attack – London, 7th July 2005 – Related Internet Sites” is well worth visiting. Interestingly, the “home”

web site of the most controversial subsequent event that occurred as a consequence of the 7 July bombings, the mistaken shooting of Jean Charles de Menezes ([www.justice4jean.com/](http://www.justice4jean.com/)), is not listed as part of the 7 July event archive. However, it is archived elsewhere on UKWAC, so that, overall, the documenting of the events of 7/7 in London is very impressive in this web archive.

Perhaps this contrast between these two examples of web archiving shows how rapidly effective collection development tools (e.g. the concept of event-based web archiving) can emerge when the LIS profession starts collection building in areas that it has never explored before. It might therefore be logical to argue that any major government report on a controversial topic should be viewed as “a web event” in itself and a host of related oppositional web sites archived as a matter of course in the appropriate national digital deposit collection.

The important thing is to start the practical business of digital collection building with a view to creating a representative national archive, without worrying that a complete model of “how to do it” is not yet in place. This may be a little worrying for a group as risk-averse as the LIS profession, but it is in effect what we are doing already. So we should have the courage of our convictions that this is an area in which it is worth making honest mistakes and we should proceed to create models of good practice by seeing what works and what doesn't work, thereby establishing a way forward.

### Conclusion

The investigations above are quite tentative and would be well worth further exploration in a considered study (a working title might be “Developing collection management principles for a new model of selective born-digital legal deposit”). If the Joint Information Systems Committee (JISC) are interested in funding such a project, they are more than welcome to take forward on a more substantial basis any of the suggestions made in this outline paper!

The essence of what is suggested here is quite simple: initial investigations of how we build national digital legal deposit collections have rightly emphasised preservation issues and the archiving of certain key collections (government web sites and dual print/electronic publications). As such, they do not give a complete basis for preserving a representative view of what the national web space was like in the past. In consequence, further work on what constitutes a sound set of web-specific collection development principles should be commissioned. The existing expertise of the LIS profession is there to be drawn on and provides an excellent starting point for such investigation.

### References

- Dictionary of Slang (1996–2006), available at: [www.peevish.co.uk/slang](http://www.peevish.co.uk/slang); version retrieved by keywords “Hutton report”, available at: [www.webarchive.org.uk/tep/11182.html](http://www.webarchive.org.uk/tep/11182.html).
- Digital Preservation Coalition (2001), available at: [www.dpconline.org/graphics/index.html](http://www.dpconline.org/graphics/index.html) (accessed 1 June 2006).
- The Hutton Inquiry (2004), available at: [www.the-hutton-inquiry.org.uk](http://www.the-hutton-inquiry.org.uk) (accessed 1 June 2006).
- Jean Charles de Menezes Family Campaign (2006), as archived on UKWAC, available at: [www.webarchive.org.uk/tep/12553.html](http://www.webarchive.org.uk/tep/12553.html) (accessed 1 June 2006).
- Legal Deposit Libraries Act (2003), Chapter 28, HMSO, London, available at: [www.opsi.gov.uk/acts/acts2003/20030028.htm](http://www.opsi.gov.uk/acts/acts2003/20030028.htm) (accessed 1 June 2006).

- 
- LIFE (2006), Life Cycle Information for E-Literature, available at: [www.ucl.ac.uk/life/lifeproject/](http://www.ucl.ac.uk/life/lifeproject/) (accessed 1 June 2006).
- Lord, H.R. (2004), *Return to an address of the Honourable the House of Commons dated 28th January 2004 for the report of the inquiry into the circumstances surrounding the death of Dr David Kelly C.M.G.*, The Stationery Office, London, HC 247.
- The National Archives (2006), UK, available at: [www.nationalarchives.gov.uk/electronicrecords/seamless\\_flow/](http://www.nationalarchives.gov.uk/electronicrecords/seamless_flow/) (accessed 1 June 2006).
- The National Archives (1998), Acquisition Policy, UK, available at: [www.nationalarchives.gov.uk/about/selection.htm](http://www.nationalarchives.gov.uk/about/selection.htm) (accessed 1 June 2006).
- Steenbakkens, J.F. (2004), "Treasuring the digital records of science: archiving e-journals at the Koninklijke Bibliotheek", *RLG DigiNews*, Vol. 8 No. 2, available at: [www.rlg.org](http://www.rlg.org) (accessed 1 June 2006).
- UK Polling Report (2004), as archived on UKWAC, available at: [www.webarchive.org.uk/pan/11808/20050716/pollingreport.co.uk/blog/index7ea6.html?p=46](http://www.webarchive.org.uk/pan/11808/20050716/pollingreport.co.uk/blog/index7ea6.html?p=46) and [www.webarchive.org.uk/pan/11808/20050716/pollingreport.co.uk/blog/indexaf05.html](http://www.webarchive.org.uk/pan/11808/20050716/pollingreport.co.uk/blog/indexaf05.html) (both accessed 1 June 2006).
- UK Polling Report (2006), available at: <http://ukpollingreport.co.uk/blog/index.php> (accessed 1 June 2006).
- UKWAC (2006), UK Web Archiving Consortium, available at: [www.webarchive.org.uk/](http://www.webarchive.org.uk/) (accessed 1 June 2006).
- UKWAC (2005), "Terrorist attack – London, 7th July 2005 – Related Internet Sites", UK Web Archiving Consortium, available at: [www.webarchive.org.uk/col/c8125.html](http://www.webarchive.org.uk/col/c8125.html) (accessed 1 June 2006).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.