

The Legal Deposit of Digital Spatial Data in the United Kingdom

by CHRISTOPHER FLEET

“Preserving digital information is becoming an increasingly urgent challenge for both libraries and publishers of books and journals, as the amount of digital information is growing quickly and preservation policies and techniques for this format remain unsettled. The need is pressing. While the costs of long-term archiving are high, the cost of doing nothing would be disastrous.” (IFLA/IPA Steering Group, 2002).

INTRODUCTION

Despite its importance for future generations, the last few years have seen only modest progress in obtaining digital data in UK legal deposit libraries (LDLs). Much has been discussed, and basic systems and agreements have been established, but progress has been limited by a number of political and technical problems. This paper attempts to set out what progress has been made and the various difficulties, from the perspective of the National Library of Scotland's involvement in the work.

UK CODE OF PRACTICE FOR VOLUNTARY DEPOSIT OF DIGITAL DATA

The United Kingdom continues to lag behind many European countries, which during the 1990s enacted legislation for the deposit of non-print publications (Denmark 1999, France 1994, Latvia 1997, Norway 1990, Sweden 1993. See: Millea, 2001; Fleet, 1999). Although some of these agreements were voluntary (as in the Netherlands), and if compulsory often only covering fixed form or offline media rather than dynamic online datasets, infrastructure and mechanisms were usually established to manage and archive new digital publications. From the mid-1990s the UK's Working Party on Legal Deposit, under the Chairmanship of Sir Anthony Kenny, made recommendations to the Department of National Heritage and subsequently to the Department for Culture Media and Sport (DCMS). However, it was not until September 1999 that the Code of practice for the voluntary deposit of non-print publications was finalised, which came into effect from 4 January 2000.

As the name implies, the Code of practice is voluntary and there is no legal obligation on publishers to comply with it. It only covers microform and offline electronic products, which are primarily text-based, or which are intended as information rather than entertainment products, and in addition, there are further exclusions. Deposit is not required if publication substantially duplicates the content of a print publication from the same publisher already deposited (which applies to several large cartographic publishers), if the publication is published for private internal use, or if it falls into various not-for-deposit categories, such as computer software, games, or film and video. It only covers publications distributed in the United Kingdom, or those first published in the United Kingdom.

The decision to deposit one copy for the British Library or multiple copies for the other legal deposit libraries is left with the publisher, although the National Libraries of Wales and Scotland can and have requested additional copies of items with significant Welsh/Scottish content/relevance. The result has been that the British Library (BL) has received the largest share of publications. For the first year (January to December 2000) the British Library received 365 offline monograph electronic publications (mainly CD-ROMs) and 235 serial titles, comprising 750 issues. The Agent for the Copyright Libraries received 342 items on CD-ROM or floppy disk (Joint Committee, 2001). By August 2002, the British Library reported that over 1,000 monographs and 850 serial titles had been received under the Code (Byford, 2002). Within the National Library of Scotland (NLS), there have been about 250 items received per year under the scheme, with a total of 539 items received by March 2002. With no national bibliography or centralised recording of electronic publications, it is difficult to know what proportions these represent of the total output, but the general consensus is that it is a relatively low one (Joint Committee, 2002).

Given the differences between publishers' and libraries' objectives, there have been several other problems. The publishers have been keen to ensure exemptions and exceptions, particularly for high-value items, and to highlight the effects of deposit on their commercial objectives. Whilst the original code allowed printing of electronic publications only to the maximum permitted under fair-dealing legislation (e.g. one chapter or one journal article and less than 5% of an item), and specifically excluded electronic downloading and saving, the publishers were keen to tighten this up, or negotiate arrangements on a title-by-title basis. Although all publishers were encouraged to supply metadata about their publications, on specially designed forms (see APPENDIX) relatively few of these forms have been filled in, with the result that library staff have had to do more work in processing items. With some justification, the publishers have also been concerned about the libraries' long-term strategies for archiving their publications, and their proposals for a secure network between the libraries. While the publishers are well-represented in the Joint Committee on Voluntary Deposit, with members from the Publishers' Association, the Association of Learned and Professional Society Publishers, the Periodical Publishers Association, and the Directory

Publishers Association, sub-groups have been set up to investigate these problems in order to make progress.

In order that the Code allows for material to be deposited in one institution, and networked securely to other legal deposit libraries, there has been much ongoing work on building a secure network between the libraries. The proposed network would be based on a distributed architecture, with multiple digital stores, and use thin-client technology (Citrix MetaFrame and Microsoft Internet Explorer) to reduce volumes of data being transferred. Unfortunately, despite a well-prepared case, both bids to the Treasury's *Invest to Save* budget in 2000-2001 were unsuccessful, and since then work has been ongoing on a "proof of concept project" base. A practical mechanism for networking electronic items therefore still looks some way away from implementation.

There has also been ongoing work on the Digital Library System (DLS) to store, preserve and retrieve digital publications within the British Library. In the autumn of 2000, the BL signed a contract with IBM for the supply and development of this system, to be based on standard IBM hardware and software and using the Open Archive Information System (OAIS) model. There has been much work on metadata to be included in such as system, and discussions with other national systems, but uncertainties and problems delay its development.

Given the limitations of the voluntary code, and the belief that its purpose was primarily a pilot, there have also been continuing attempts to introduce legislation for the compulsory deposit of electronic publications. The DCMS put forward proposals for such legislation in November 2000, but the bid was not successful. From 1998, such new legislation requires a Regulatory Impact Assessment (RIA), noting in particular the costs of compliance to business, and publishers did not approve the BL's provisional Assessment. Since then there has been more extensive work, and a contract was awarded in 2002 to Electronic Publishing Services Ltd. to conduct a full RIA. In the autumn of 2002 there have been efforts to introduce new legislation through a 'handout bill' (in effect, a government-sponsored private-members' bill). Even if such proposals are successful, it's unlikely to take effect before 2004 (Bury, 2002).

Within the NLS Map Library, it is difficult to claim that more than a handful of electronic cartographic products have been received as a result of the Code. Of course, many cartographic publishers maintain and sell online datasets (not covered by the code) or supply paper print publications (Hydrographic Office, Automobile Association, Ordnance Survey small-medium scale mapping) in lieu of electronic products. Given the lack of expertise within the NLS over archiving electronic publications, the relatively low quantities of incoming electronic items is to be welcomed, and it is for this reason that the NLS still prefers to acquire conventional paper mapping over its electronic equivalents. However, such a situation undoubtedly means that some current electronic products are being lost to future generations.

MAP LIBRARIES' NEGOTIATIONS

Given the delays and limitations of the voluntary code, the six United Kingdom and Ireland LDL map libraries (Bodleian Library Oxford, British Library, Cambridge University Library, the National Libraries of Scotland and Wales, and Trinity College Dublin) have actively sought agreements for the supply of cartographic digital data. These relate to three main publishers: Ordnance Survey of Great Britain (OSGB), Ordnance Survey of Northern Ireland (OSNI), and Experian Goad. Whilst there has been considerable progress with OSGB data, there has, regrettably, been very little progress on OSNI and Experian Goad.

1. Ordnance Survey of Great Britain (OSGB)

As described in Fleet (1999), there were long and protracted negotiations with OSGB over the supply of their digital data to legal deposit libraries. There were also several technical issues to resolve, such as the customisation of software to view the Ordnance Survey (OS) data, the need to archive and convert the Land-Line data, and the need to agree developments collectively between libraries, OS and the University of London Computer Centre (ULCC), the BL's preferred electronic archive. Nevertheless, the libraries were very grateful that OS was prepared to send their data to the LDLs, albeit with strictly controlled usage conditions and security agreements, given the absence of compulsory deposit legislation. The succeeding problems and delays in implementing the system have been primarily due to difficulties within libraries, and not to OS.

On the positive side, there has been progress on a number of fronts. Whilst the National Library of Scotland has run a pilot viewer system for the public from 1998, the other libraries have subsequently made progress, and the OS Viewer System has been set up to varying degrees in all other libraries during 2002. Several design modifications (tweaking certain functions) were agreed between the LDLs during 1999 and successfully implemented through Dotted Eyes, also (in the process) correcting several software bugs and problems. Although the delays to signing a security agreement between OS, BL and ULCC until August 2000, which resulted in a backlog of annual snapshots, progress during 2001-2002 has virtually cleared the backlogs, and LandLine data for 1998, 1999, 2000, and 2001 has all been converted and passed on to the LDLs. As the original agreement covered just LandLine topographic data (with no contours and limited height information), the OS agreed in August 2000 to supply LandForm profile data to libraries at no additional charge. During 2001 there was progress in re-customising the Viewer to incorporate and display this height information. Finally, the various security agreements between the LDLs and OS were finalised during 2000-2001, based on the BL/OS agreement of September 1999, the Viewer was licensed for the LDLs to use, and a support agreement was drawn up for the maintenance and troubleshooting of the software.

However, the progress has not been smooth or speedy, due to three related factors:

1. Political/resource problems

The speed of developments have been influenced by the degree to which the British Library, specifically the Information Systems Department (IS), could devote time to OS digital data. As founders and owners of the Viewing software, as well as managers of the ULCC conversion and archiving, collective progress hinged on the BL. Other priorities within the BL, including reorganisation and job cuts as well as high turnover of staff within the IS meant that little progress was made at all during 1999 and 2000. In a couple of instances, BL IS staff moved on not long after gaining the required knowledge of the system. Following repeated requests from the LDL map libraries (as well as BL Map Library) it was only through official pressure through our librarians/chief executives that progress was eventually made at BL during 2001-2002.

Part of the problem also relates to the degree to which map libraries are seen as fringe concerns within their host institution, and indeed, the degree to which the libraries themselves are seen as a fringe concern to Ordnance Survey. The former problem varies between the LDLs, but within the NLS it has certainly caused difficulties in getting funding for hardware/software, and technical support for set up and maintenance. The latter problem is debatable, but in an online environment, libraries are tending to play a more peripheral role in distributing OS mapping information, and the demise of OS Consultative Committees in 2001 has denied BRICMICS its direct channel of communication and influence with OS.

The OS digital data exposed some of the contrasts between the LDLs. The NLS position has tended to see the OS digital data as a central part of its operations, with high usage by a range of present-day users, and therefore an impatience to implement the system as soon as possible. On the other hand, the BL position has perhaps seen it as a less central part of their present Map Library services, of greater interest and value in the long-term as an archive for historical research, and they have consequently been happier with a longer time-scale of implementation.

The net result of these factors is that the libraries have moved at a much slower pace than Ordnance Survey, and are actively implementing a system with data that we know will be superseded within the next few years. First, there were developments in LandLine in 1999 to allow the date stamping of feature codes. This information was considered to be of great value for future historians of the landscape and the LDLs, but would have required re-written conversion software to utilise. Given the problems mentioned above in getting so many more fundamental agreements and systems sorted out, this was not pursued, and the data is therefore not in the LDLs' systems. Second, and of greater importance, however, has been the development of the Digital National Framework during 2000-2001, a process whereby OS have re-engineered their entire large-scale database on a feature-based, rather than tile-based system. Marketed now as MasterMap,

the new data allows much greater querying, linking, analysis, and real-world representation than the former LandLine system. However, its fundamental differences from LandLine mean that historical LandLine data cannot be migrated into the new format. Also integrating the two data formats for date comparisons within a single system would be very difficult, and even methods of quantifying change across data formats may need to alter. Whilst OS have promised to support LandLine for a few years until their customers have converted to using the new data, the LDLs have not yet started discussions over this transition. We have also not decided whether to integrate LandLine and MasterMap in one application, or keep them separate.

2. Technical problems

Given the use of specialised and customised geographic software, and the need to convert, distribute and query a body of over 100 Gb of data (for three annual snapshots alone), technical problems were inevitable. Yet, arguably, these were not as significant as the political ones in hampering progress. The requirements placed by OS on what we could do with the data, and the lack of any suitable off-the-shelf product demanded that software was customised. Not only was this relatively expensive, with the need for maintenance and support from the Dotted Eyes, but we have also been somewhat tied to Dotted Eyes for future modifications and enhancements. It has also meant that relatively few people have much knowledge about the software, and even experienced IT staff takes time to gain familiarity with the system and its background. Real progress in sorting out systems at the BL in 2001-2002 has only happened with staff with a genuine desire to understand the system and sort out problems.

There have also been some difficulties in converting, transferring, and loading data through ULCC. Again, some of these have been due to the specialised nature of geographic data, whilst others have related to the different operating system used at ULCC (UNIX rather than Windows as in the LDLs), problems in reading DAT tapes, and difficulties in loading large volumes of data and the metadata related to it. (A typical annual snapshot can consist of over 220,000 tiles, which translates to nearly 900,000 MapInfo files). It is relevant to note that it was only through the ULCC noting data corruption in the 2001 snapshot from OS, that OS were made aware of this problem within their own archive.

3. Costs

Although OS have not charged the libraries for the data itself, there are significant costs in converting, archiving and distributing the data, as well as supporting the Viewer:

<u>Cost of Annual OS update</u>	
Conversion of annual update	£2,400
Archival storage of source data*	£1,000
Archival storage of converted data*	£1,000
Supply of data to the LDLs	£ 700
Viewer support contract	£4,600
Total	£9,700
Cost per LDL	£1,600

In addition, the * items are cumulative annual costs that will grow on an annual basis along with the data. There are also ongoing special costs, such as the recustomisation of the Viewer to take PROFILE contours, which in 2001 cost ca. £8,000. To these costs, which fortunately are shared centrally between all 6 LDL libraries, are the costs for each library of appropriate hardware/software, and considerable staff time to convert data, training, and maintenance. Whilst the total annual costs have so far tended to be lower than the typical annual cost of mounting the SIM microfilms during the 1990s, the cumulative cost of digital data is significant. Even allowing for a conservative estimate of these costs, and the fact that a different process and archive might incur different charges, OS digital data is expensive, with steadily growing costs of archiving over time.

2. Ordnance Survey of Northern Ireland (OSNI)

As with OSGB, OSNI have moved to producing digital data instead of large-scale paper plans, and the LDLs were also interested (to differing degrees) in obtaining this digital data. Given the similarities in format (both used a tile-based National Transfer Format) and concerns over use, it was suggested to OSNI that the LDLs might use the OSGB arrangements and security agreements as a suitable template for receiving their data. With the delays in finalising these agreements, this proposal could not be put formally to OSNI until 2001, but they responded very positive. In principle, they were willing to supply their data to the libraries, and were keen to allow sample data to be tested. Unfortunately, the practical investigation of the data and conversion parameters, as well as tweaking the Viewer to use their data required central BL IS staff time, which has not been forthcoming, putting progress on hold during 2002.

3 *Experian Goad*

Large-scale Goad Fire Insurance Plans date back to the late 19th century for several British cities, and provide unique information on business premises, retail outlets, and industrial units through time. Many LDLs received these comprehensively from the late 1960s, with updates every year or two years, for ca. 1,200 UK city centres. Although the maps were supplied under legal deposit as (light-sensitive) dyeline prints, and therefore of low archival stability, the plans were a useful, more frequently updated addition to Ordnance Survey large-scale mapping for town centres. However, in 1998 Experian Goad informed the Copyright Libraries' Agent that as the plans were produced digitally, as printouts on demand, they were not conventionally published as copyright maps and therefore should not be supplied to the LDLs.

In 1999-2000 the LDLs response was to focus on getting OSGB data sorted out, with the hope of arranging something suitable with Experian Goad, such as obtaining digital Goad plans. The plans themselves can be read through MapInfo, the software used for viewing the OS data. Although there was enhanced querying and customisation facilities with the digital format, there were substantial costs of acquiring the data, and royalties for printouts. As no progress was made, and the Goad plans were not supplied to the LDLs from 1999, one LDL mentioned in 2001 that they had acquired a subset of Goad plans to at least continue their paper archive. Amongst other things this highlighted that the plans themselves did seem to carry a similar status to conventional paper publications, and therefore a formal letter was sent via the Copyright Agent in May 2002 to request again that Experian Goad deposit these plans.

CONCLUSIONS

With publishers' continuing and growing concerns over use of their data in libraries, which led to a reduction of their revenues, relatively few high-value items are being sent to libraries. It seems that real progress in digital deposit will only be made through compulsory legislation and/or with negotiated arrangements over usage and access. The following general conclusions can also be made:

- Given the lack of progress at both national library and map library levels, electronic cartographic publications in the UK are not being comprehensively acquired, made available to the public, or archived.
- Progress for map libraries has only really been made when people, with time and reasonable IT proficiency have been given responsibility for the work.
- Cartographic digital data is changing in format, media, and content faster than the LDLs can currently keep pace with.

The Legal Deposit of Digital Spatial Data in the United Kingdom

- For cartographic publications, particularly those of high commercial value, direct negotiations between libraries and publishers may be essential in agreeing specific access and usage arrangements.
- Digital archiving is relatively expensive and complicated, especially when using non-standard software and specialised geographic data.

Whilst the LDLs have co-operated well and have shared certain costs, their conflicting priorities, inadequate distributed expertise, and the expense of duplicated systems within the LDLs have all caused problems. From the perspective of publishers and the BL, a centralised online supplier, along the lines of EDINA or MIMAS, would arguably be a more cost-effective, efficient way of managing electronic publications, able to keep pace with technology, and easier to manage.

REFERENCES

1. Bury, L.: "Digital deposit costs assessed". *The Bookseller* 26 July 2002
2. Byford, J.: "Publishers and legal deposit libraries co-operation in the United Kingdom since 1610: effective or not?" Paper read at 68th IFLA Conference, Glasgow August 2002. <http://www.ifla.org/IV/ifla68/papers/126-140e.pdf>
3. Fleet, C.: "Ordnance Survey digital data in UK Legal Deposit Libraries". *LIBER Quarterly* 9 (1999), 235-243. <http://www.kb.nl/infolev/liber/articles/fleet11.htm>
4. Fleet, C.: "Map Curators and the European Context". *Proceedings of the British Cartographic Society 36th Annual Symposium* (Glasgow, 1999), 10-15.
5. Fraser, C.L.: "Closing the deposit gap". *The Bookseller* 6 July 2001, 27-29.
6. HMSO Guidance Notes. *The National Published Archive - Legal Deposit of Official Publications. Number: 11 Date: 15 May 2000 (Revised 6 November 2000)*. <http://www.hmso.gov.uk/g-note11.htm>
7. IFLA/IPA Steering Group. *Preserving the memory of the world in perpetuity: a joint statement on the archiving and preserving of digital information*. 27th June 2002 www.ifla.org/V/press/ifla-ipa02.htm
8. Joint Committee on Voluntary Deposit. Annual Progress Report, 2001
9. Joint Committee on Voluntary Deposit. Annual Progress Report, 2002
10. Millea, N.: "Organisational Change". In: *The map library in the new millennium*, ed. by R.B. Parry & C.R. Perkins. London : Library Association, 2001. Pp. 28-42.

WEB SITES REFERRED TO IN THE TEXT

Code of practice for the voluntary deposit of non-print publications.

<http://www.nls.uk/professional/legaldeposit/nonprint/code.html>

Edinburgh Data and Information Access: <http://edina.ac.uk>

Manchester Information & Associated Services: <http://www.mimas.ac.uk>

APPENDIX

*Code of practice for the voluntary deposit of non-print publications: Form 2:
Publication Specific Information*

To assist in processing of the publication please complete and send a copy of this form with each publication deposited.

1 Bibliographic information

- Title of publication
- Author /creator of publication (if appropriate)
- Frequency (If serial)
- Volume / part number.
- Standard number
- Publisher
- Place of publication
- Year of publication
- Medium/format: Microform: 16 mm roll / 35 mm roll / Microfiche

Offline electronic: CD ROM / DVD / Magnetic Disk / Other (please specify)

2 Technical information

Please provide on separate sheet(s) or by attachment of documentation the technical information needed to use the publication under the following headings:

- Hardware requirements. (Describe both minimum and optimal hardware platform needed)
- Operating system requirements. (Include version number, language and locality)

The Legal Deposit of Digital Spatial Data in the United Kingdom

- Associated software requirements. (Describe any other software needed to use the publication)
- Installation information. (Describe any settings or other information needed to install the publication)
- Format of content

Please enclose any additional technical information needed to process, use or make a preservation copy of the publication (see also 4 below).

3 Access arrangements for offline electronic publications (only if different from previously specified)

- Are you willing to deposit copies of this publication to each of the six legal deposit libraries? Yes / No
- If 'Yes' , please specify the access arrangement permitted within each holding library. (Please tick one box)
 - a) Single user at a time via an internal network (default option)
 - b) Single user at a standalone workstation
- If you are willing to deposit copies of this publication only to a single legal deposit library, please specify the access arrangements permitted. (Please tick one box)
 - a) Single user at a time via an internal network within the holding library (default option)
 - b) Networked access between the legal deposit libraries to a single user at a time across the whole network
 - c) Networked access between the legal deposit libraries to a single at a time in each library
 - d) Single user at a standalone workstation within the holding library

4. Copying of deposited electronic publications for preservation purposes

It is assumed that copying of the publication onto another medium for preservation purposes only is permitted, subject to the preservation of the individual publication's identity and integrity. The copied version will not be used to provide user access. Please tick this box only if you do not permit this copying for preservation purposes

5. Contact information in case of queries

Name:

Organisation:

Phone:

E-mail: