

This is a preprint of a chapter accepted for publication by Facet Publishing. This extract has been taken from the author's original manuscript and has not been edited. The definitive version of this piece may be found in 'Electronic Legal Deposit, Shaping the Library Collections of the Future,' Facet, London (ISBN 9781783303779), which can be purchased from <http://www.facetpublishing.co.uk/title.php?id=303779>. The author agrees not to update the preprint or replace it with the published version of the chapter.

Our titles have wide appeal across the UK and internationally and we are keen to see our authors content translated into foreign languages and welcome requests from publishers. World rights for translation are available for many of our titles. To date our books have been translated into over 25 languages.

Making History:

William Kilbride, Digital Preservation Coalition

It is hard to imagine a more interesting time to work in libraries, nor a more challenging one. In an era of post-truth obfuscation and sinister deletion, the ability to collect, retain and authenticate is suddenly a super-power; in an era of relentless proliferation, the confidence to select and consolidate, with implied permission to relegate and de-duplicate, is ubiquitously essential; in an era where data is the 'new oil' of the 'information society', the unassuming librarian holds the keys not only to the past, but now also to the future. One would have thought that this generation more than any other would be the age of the library, an enduring proof of common cause for the common weal: deposit libraries at the summit of our ambition, the record of all we have achieved and source of all we might. Why does it not feel that way?

It's not yet clear whether the digital turn will be the making of the library or its undoing, given many of these opportunities are disruptive, mostly provisional, and largely originate outside of the library community. These challenges arise just at a moment where the social and economic context of operations are profoundly unsettled, whether through the continuing dysfunction of economics, the puzzling impasses of public discourse, or a global crisis of dislocation and dispossession. With such uncertainty about the times in which we will shortly live, this is no time for an identity crisis. Yet there is little prospect of staying unchanged.

2018 has been (another) pivotal year in the development of libraries in the United Kingdom and Ireland, a consequence of a review of the legal deposit regime. This has enabled a wider consideration of the place and function of legal deposit libraries, and a deeper dive

into the evolving processes and expectations through which their mission, to provide a full and canonical copy of the published record, is fulfilled. So how can libraries help us make sense of these turbulent times and preserve a record that will help us look back on them too? How can we ensure that legal deposit libraries and their collections will continue to thrive in the second quarter of the twenty-first century?

The *Legal Deposit Libraries Act 2003* made explicit for the first time in the UK provisions for publishers to deposit on and off-line publications and about the use and preservation of such publications through by the legal deposit libraries. Further regulations published in 2013 implemented the 2003 Act with specific guidance around copying and adapting materials for the purposes of preservation (HM Government 2013 and DCMS 2013). Five years on from this secondary legislation, the UK Government Department for Digital Culture Media and Sport has initiated a review of the regime. This has provided an opportunity to assess how successfully the legal deposit libraries have implemented all aspects of the regulations and it has inevitably provoked informal discussion on the extent to which the regulations are fit for purpose. In preparation for the review, the legal deposit libraries invited the Digital Preservation Coalition (DPC) to undertake a specific and independent assessment of digital preservation capability for non-print legal deposit collections. This was undertaken throughout 2017 and a full report was published in March 2018 (see DPC 2018). From the perspective of posterity, this assessment is an important measure of the continuing value and presumed role of legal deposit libraries in the future.

This chapter will review five aspects of electronic legal deposit drawing from the lessons learned in the DPC's assessment of digital preservation capability. It will describe how the legal deposit libraries in the UK and Ireland have addressed the digital preservation challenge; it will summarise the findings of the assessment, identifying the main issues and challenges which the legal deposit libraries have faced; it will explore whether the experience gained so far has prepared them well for the future; it will describe a emergent challenges associated with users in which the regulations are seen to be at odds with standards in digital preservation and significantly adrift from current theory in memory institutions; and it will describe an emergent challenge with technology where the regulations assume a false dichotomy between software and data which will become

unsustainable. In doing so, this chapter will propose renewal and enhancement to the mission of legal deposit libraries, preparing them for the challenges to come and glimpsing a prospect that may transform them from problem owners uncertain about their role, to solution providers that are confident about their purpose.

1. How have legal deposit libraries addressed the digital preservation challenge?

Libraries have been among the first and most committed champions of digital preservation: and by extension early and steadfast adversaries of obsolescence (see for example Hedstrom 1998, Brindley 2000, Shenton 2000, Webb 2000, Day 2001). Many of the tools, processes and models of the digital preservation community were developed specifically to support library applications or have improved and advanced beyond recognition when adopted by them. For example many of the tools and approaches of digital forensics were developed explicitly in the context of legal processes but have found a significant and alternative applications in the stabilisation and assessment of digital manuscript collections (John 2012); many of the technologies of web publishing have been adapted to create robust new methodologies for web archiving (Pennock 2013); and the out-sourced preservation of e-journal content in the context of library collections is in some ways a model for all digital preservation architecture (Beagrie 2013). Consequently, legal deposit collections are well served with digital preservation capability relative to other sectors and collection types. However, digital preservation is a subtle challenge, emerging from the commercial and technological caprice of social, industrial and economic forces outside of libraries' control: what was once good practice may now be insufficient, and what is now robust may soon become uncertain. For example, electronic journals still deploy paper metaphors of files and print but when publishers drop those metaphors and produce truly electronic journals that are complex multi-media interactions of data and syntheses, then a new preservation challenge emerges: in fact, it is already here (Kirchhoff and Morrissey 2014, Day et al 2018).

The responsibility to care for digital materials is absorbed within a broader social and cultural mission of legal deposit libraries: to 'make a significant and lasting contribution to global knowledge and the memory of the world' (National Library of Scotland 2015, 5) whilst ensuring that we can 'guarantee access for future generations' (British Library 2010, 3).

These responsibilities are described more prosaically, in the regulations that specify their

rights and mandates that shape their processes. In the UK and to a lesser extent also in Ireland this means the 'UK Legal Deposit Libraries (Non-Print Works) Regulations 2013' which implement the UK 'Legal Deposit Libraries Act 2003' which, *inter alia*, makes provision for the use and preservation of material deposited for what is termed 'Non-Print Legal Deposit' Collections.

In practical terms implementation of the legal deposit regulations are overseen by the Joint Committee for Legal Deposit, composed of representatives of the publishers and the legal deposit libraries¹. This committee receives reports on the implementation of the regulations in order that it can evaluate the effectiveness of non-print legal deposit arrangements, assess the application of the Regulations in practice and future needs. The committee has gathered evidence in preparation for the UK Government's review of the regulations, taking into account the views of specialist groups such as the DPC. The committee convenes a number of task forces and sub-committees from time to time, including a group for overall implementation of the regulations which has digital preservation among its responsibilities.

There are six library partners involved in the legal deposit regime in the UK and Ireland, and each of these have significant digital preservation expertise (for recent and relevant examples, see the National Library of Scotland's work on cloud storage (Hibberd 2017 and references) or the British Library's work on preservation of diverse media formats (Day et al 2016)). Digital preservation for non-print legal deposit collections is delivered by the British Library who are effectively service providers to the other five deposit libraries. This concentration provides economy of scope against a single shared set of challenges. It also means that, of all the different aspect of the legal deposit regime, digital preservation ought to be among the simplest to assess. One question can be posed to one institution: are the digital preservation skills, policies and processes of the British Library sufficient to ensure the non-print legal deposit collections will be accessible for the long term?

¹ For information including terms of reference and membership of this committee see: <https://www.bl.uk/legal-deposit/joint-committee>

2. Assessment: Methodology Scope and Findings

There have been numerous attempts in recent years to codify good practice in digital preservation into a series of standards (see Bantin et al 2016 for an introduction). The idea of a trusted digital archives was first expressed in 1996 by a joint working party of the Commission for Preservation and Access and the Research Libraries Group which called for independently-administered program for archival certification (CPA/RLG 1996, 9). The concept of the 'Trusted Digital Repository' emerged from that as a core theme at around the same time as the legal deposit regime was expanded to encompass non-print collections (RLG/OCLC 2002). This has since been developed in a series of related audit and certification standards over the years (e.g. CRL 2007, ISO 2012).

Given these existing tools, the assessment of what makes a Trusted Digital Repository should be easy to scope, but there is a subtle tension between the concept of the 'trusted repository' as a model of good practice and the force of quasi-judicial regulations. The mismatch between the certification standards in digital preservation, the context of preservation services and the nature of technological change has been reported before (Rosenthal 2014, Kilbride 2017b, Duranti 2016). The public purse is not funding a trusted repository as defined by some abstract standard, but a trusted series of tools, services, procedures and staff, configured around an interpretation of statute and regulation. For example, questions of mandate which are explored at length in the context of the trusted repository audit (such as sections 3.1 3.2, 3.3 and 3.4 of ISO 16363), do not arise, and many aspects of security, staffing and sustainability fall out of scope. In this context the standard metrics of digital preservation capability are more usefully seen as a toolkit in the development of an assessment. Because such organizational metrics were redundant, and because an internal audit using ISO16363 had already been completed (Pennock and Smith 2016) the opportunity arose for a deeper analysis of operational and governance matters as they related to specific collections.

The details of the resulting *Non-Print Legal Deposit Digital Preservation Review*, including the metrics, processes and findings have been published in detail (see DPC 2017 for a complete account). They are summarised here to provide a flavour of the challenges and competences which frame the subsequent discussion.

As well as adapting metrics from the Unified Requirements for Core Certification of Trustworthy Data Repositories (The Core Trust Seal) the review had the luxury of ensuring greater independence, transparency and penetration than might normally be delivered. The DPC was commissioned to examine practice at the British Library by the Joint Committee on Legal Deposit and ultimately reporting to and accountable to them. Although the British Library is represented on this committee, so are all the other legal deposit libraries and significant representatives of the publishing trade. This guaranteed the independence of the assessment, in contrast to conventional processes for the Core Trust Seal which is normally awarded after lightweight peer review, carried out by other holders of the Core Trust Seal. The DPC acted as an independent reviewer, requesting information and interviewing staff across legal deposit institutions and in particular the British Library to ensure consistency and probe strengths and weaknesses. This process is more exacting, enables a more forensic and detailed analysis of strengths and weakness, and ensures that there is no inadvertent misunderstanding. It is also more transparent since there is no conflict of interests between assessor and candidate. The DPC, as an independent not for profit which is owned by its members and supports them delivery the strategic purpose of a secure digital legacy has engrained expertise in the topic but will not in turn be subjected to the same analysis, thus obtaining no benefit from eroding or amplifying the metrics in its favour.

The assessment included processes, data, staffing, skills, planning and policy (see DPC 2017). Each of the legal deposit libraries were invited to participate and receive reports on progress at key stages in the process. Moreover, the assessment was structured in such a way that it could be repeated, whether by the legal deposit libraries at some later date, or by other institutions seeking validation of their digital preservation activities (see DPC 2017 Appendix One for a list of those interviewed and Appendix Two for the metrics used). It examined three primary non-print legal deposit collections which constitute the vast majority of electronic legal deposit: web archives including both the domain crawl and voluntary deposit; non-print deposit of journals; and non-print deposit of e-books. The review also examined progress with two other content types which are in scope of the

regulations but are relatively small in comparison: digital manuscripts of music, and cartographic data.

The review progressed iteratively starting in January 2017 and completing in December 2017. The process consisted of five phases: a period of preparation and initiation; an early assessment phase and report; a period in which recommendations could be acted upon; then a second assessment phase which examined the extent to which recommendations had been followed; and a reporting phase where all the documents and outcomes were shared and published. It focused on the current state of the preservation capability of non-print legal deposit materials in the UK and Ireland, but also took the opportunity to consider how this capability has changed since the introduction of the Legal Deposit Regulations in 2013, and where current plans may take it by 2023. In this way it assessed and demonstrated responsiveness to a 'moving target'.

Significant conclusions from this assessment can be summarized in six key points.

First, it is hard to express the scale and complexity of the challenge which the preservation of digital legal deposit collections has generated. Moreover, it is a constantly changing challenge. In the face of this expansive, challenging and changing workload there is a constant need for renewal of resources, skills and technology. In less than five years the non-print legal deposit collection at the British Library has grown from 365,000 to almost 14,000,000 digital objects, a count which if anything underestimates the abundance of web content since entire collections of websites are managed as composite WARC file containers². This *thirty-eight-fold* increase has not been met with an equivalent increase in resources.

Secondly, in order to make progress with this exceptional growth and complexity the British Library has made sensible, clear-cut strategic decisions with respect to the management of resources and skills. Larger content streams, that is web archives, e-journals and e-books,

² See Pennock 2013 17 for an account of the WARC file and its practical deployment

were prioritized in the early phases of implementation and focus has since moved to the smaller but more complex and interactive streams which are a greater technical challenge.

Thirdly, a number of operational failings were observed in the review such as technical difficulties with integrity checking and backlogs in accession and replication. Integrity checks are a vital component of digital preservation processes, establishing that the materials currently stored are an exact match to those initially received. These can be time consuming to process and therefore question arises as to how frequently such checks should be processed and how long it is necessary to maintain a log of such audits. New non-print legal deposit collections arriving in the British Library were not immediately ingested to the digital library system where they would be automatically replicated to the other Legal Deposit Libraries. A short delay for validation and quality assurance is good practice but lengthy delays mean risks to the content are not fully managed. Challenges in both areas were noted, and in turn revealed difficulties in escalation and oversight. These were reported in the course of the assessment, and subsequent checks verified that actions had been initiated to resolve them. In most cases these were complete by the time of the second review.

Fourthly, these technical challenges turned out to be relatively quick to address: discussion in the preliminary phases made failings and solutions obvious meaning many were resolved even before they were reported. For example, the Object Authenticity Checker was completely re-written in the course of the project and a full integrity check was completed, exceeding the reviewers' expectations that this would be possible. The combined problems of integrity checking and backlogs in could have resulted in significant loss of NPLD content. But this new integrity check has revealed this not to be the case. Management challenges have been addressed quickly through enhanced reporting to the other legal deposit libraries, and a new specialist task group for the oversight of digital preservation challenges has been established which can report to the Joint Committee. As with many cultural changes these may take longer to become established norms. Changes to governance should be closely monitored to ensure they remain effective.

Fifthly, it is pleasing to report that the legal deposit libraries, and the British Library in particular, responded positively, effectively and in a timely manner to the *Non-Print Legal Deposit Digital Preservation Review*. The reviewers came to an independent judgement with detailed knowledge of best practice and standards around the world, and with full access to all relevant documents, processes and staff. This demonstrated a firm commitment to continuous quality improvement.

Finally, and perhaps most importantly the assessment demonstrated that much of the digital preservation practice at the British Library, on behalf of the legal deposit libraries has been exemplary. The British Library has been a world leader in many aspects of digital preservation for some time (e.g. Shenton 2000, Brindley 2000, John 2013, Pennock 2014). This expertise has been wisely and consistently brought to bear on the non-print legal deposit collections. For example, the proliferation of publisher formats creates significant challenges for characterization, the act of identifying a file as belonging to a specific format, and validation, the act of confirming that it conforms to the standard or identifying the ways that it fails to conform and the risks that arise. Not only has working practice kept pace with rapid and unpredictable variations, the British Library has engaged actively in pushing forward thoughtful understanding and development of best practice which it has shared and validated in conversation with the global digital preservation community (eg Day et al 2018). Examples like this enabled the reviewers to describe the British Library as a sector-leading digital preservation institution with global reach.

3. Future prospects: Non-Print Legal Deposit in 2023 and beyond

Six significant challenges emerged for the review which will need to be met to ensure that the legal deposit libraries remain capable to address digital preservation challenge in five years' time:

First, the assessment has measured the extent to which collections continue to expand. As publishers move their already-published holdings from print to digital, new submissions not only include current publications but also back copies originally published in print. This increase in volume makes planning acquisitions challenging. 42% more content is received

from publishers in electronic form compared to when the same publishers deposited only in print (DPC 2017 10).

Second, digital materials invite revision and correction in ways that are challenging. Revisions ranging from minor correction of punctuation through to significant edits have been requested. In such circumstance deposit and accession processes may need to be carried out several times over.

Third, publishing formats and processes are not stable. Publishers and their outputs change over time. This include business and technical changes such as mergers with other publishers or corporate entities, closing or changing the name of their journal titles. These unpredictable changes result in further complication to established content workflows.

Fourth, digital collections are liable to repackaging in ways that physical collections are not: material previously published in one form can be repackaged and published (otherwise identically) in different granular form; for example, journal articles can be re-published as an e-book.

Fifth, digital collections invite richness and interactivity. Sometimes these forms match well with emerging standard data formats, in other cases content becomes tied to proprietary software technology, presenting obstacles to preservation, management and access. Digital deposits are typically unpredictable, and the simple act of depositing the data can result in challenges itself. A publisher established as a submitter of content to the British Library for preservation typically supplies content in a gradual fashion. However, the successful completion of negotiations with new publishers means that an entire back catalogue may suddenly appear for submission. This unpredictability of the frequency and volume of deposits makes forward planning and the day-to-day management of ingest workflows a continual challenge.

Finally, increasing volume, complexity and unpredictability of content place considerable strain on digital preservation workflows. Greater volumes (both in numbers of items and sizes of component files) place strains on systems that process them, requiring more

resilient software processes and greater automation to enable issues to be resolved without backlogs arising. Backlogs, as noted elsewhere in the assessment, exacerbate risks faced by collections. Evolving complexity requires evaluation and research into new file formats, and new types of digital content. In some cases, this may require new preservation techniques and, most likely, evaluation and implementation of new software applications to handle them. Unpredictability requires greater flexibility to react to changes in content and its supply. Deposited data that doesn't conform to previously encountered norms must be detected, and workflows adapted to process it. The accuracy and completeness of digital preservation activities will be impacted with adaptation to meet these challenges.

4. Why users matter: making sense of non-print legal deposit.

The challenges of preserving non-print legal deposit collections are unpredictable. Unlike physical preservation, the decay of digital materials does not follow predictable chemical or biological vectors which the library may seek to prevent or postpone: the causes of obsolescence and media failure are ultimately typically the result of mercurial economic and business processes that are outside of the library's control (I have tried to argue this issue from a historical perspective (Kilbride 2017b), but for a more succinct technical overview of the causes of obsolescence see Harvey and Weatherburn 2018 50-52, Brown 2013 200-208). This is why digital preservation experts lay such emphasis on adaptability (Harvey and Weatherburn 2018 204-5 and references). There are two areas where the current regulations appear to constrain this essential flexibility: in relation to users, and in the paper-based definitions of collecting scope.

The 2013 Legal Deposit Regulations are quite specific and restrictive about the contexts in which collections can be accessed: materials can be consulted at a terminal on premises controlled by the deposit library, and moreover the library must ensure that only one computer terminal is available to readers to access the same relevant material at any one time. There seems little doubt that this restriction to access, modelled loosely on restrictions applied to print collections, is a significant constraint on users and the potential impact of non-print legal deposit. There is a case to be made that these constraints to access will have a deleterious effect on preservation. Conservation of physical materials benefits from restricted access and use while digital materials typically benefit from continued use and the subtle quality assurance this implies.

Digital preservation planning and quality assurance typically revolves around users expressed formally in the design and delivery of '*dissemination information packages*', '*representation information*' and the '*designated community*' (See Lavoie 2014 for an accessible explanation of these terms and ISO 2012 for explicit definitions). To summarize, representation information is a component of an archival information package which holds semantic and structural components, which in turn provide the transparency required to ensure that a 'digital object' can become an 'information object' in the hands of a 'designated community' (Gartner and Lavoie 2013). Arguably, representation information, is the unique characteristic that distinguishes digital preservation from every other kind of content management. Representation information is interpreted using representation information which, at face value, implies a sort of recursive absurdity in which each information package is required to contain or reference the entirety of human knowledge to achieve independent utility. The Open Archival Information System (OAIS) standard avoids this in two ways: by linking representation information into networks; and by accessing an underlying *knowledge base* which can be taken for granted (Lavoie 2014). For example, a person who has a knowledge base that includes an understanding of English will be able to read, and understand, an English text. So, the requirement for representation information is curtailed when it is mapped against implicit knowledge between two agents within an information exchange. Digital objects are preserved therefore through an ongoing configuration of data object, representation information and the implied knowledge of the designated community.

This is self-evidently useful in the face of the alternative infinite regression. However, it implies a significant effort to understand the needs of a designated community and an active engagement to track their changing needs. Archives which fail to do this cannot ultimately deliver long-term preservation. Reading this requirement through the lens of legal deposit regulations creates a tension, since the constraints to access which protect the copyright of publishers undermine the potential for dialogue with the designated community and thus undermines, and ultimately will defeat, the preservation mandate. In the long run preservation and access cannot be completely separated.

There is arguably a third, albeit more abstruse, challenge to be considered in the relationship between access and preservation, derived from deeper cultural and linguistic theory.

Over the last two decades, digital preservation has been configured as a collaborative endeavour that draws for many disciplines adopt tools and good practice from wherever they can be found. For example, Open Archival Information System (OAIS), which is in many senses the lingua franca of digital preservation, is the product of the Consultative Committee for Space Data Systems (CCSDSS 2012). This origin haunts the language and assumptions of digital preservation: the values and norms of space science lurk below the surface of just about every digital preservation conversation³. This contribution has been immensely welcome and potent, but it also should be set alongside contemporaneous progress in library, archival and museological theory which emphasize the difficult and at times contradictory nature of meaning-making. The textual turn of cultural hermeneutics and poststructuralism in particular have been controversial: the former situating cultural studies within wider social and political fields of discourse in which social relations are created and reproduced (see Adorno and Horkheimer 1972, Hall 1990, Spivak 1990); the latter denying the sufficiency of established approaches to meaning making in structural linguistics (Derrida 1976, Barthes 1973). It has been argued that the whole genre of post-truth informatics has some origin in the legacy of postmodernism (D'Ancona 2017). It has certainly been a mixed blessing for archives, libraries and museums.

On one hand, the recognition that knowledge production is a fundamental tool in the reproduction of unequal power relationships has transformed memory institutions from the gatekeepers of authoritative resilience to the enablers of progressive narrative(s). Derrida equated archives with a sort of house arrest: both as the source and containment of power, arranged to the practical convenience of the authorities, and only shared on asymmetrical terms with the public (Derrida and Pernowitz 1995 10). For several decades, anthropologists have traced the machinery of knowledge production and maintenance in explicitly colonial

³ The full title of the Open Archival Information System standard is: *Space data and information transfer systems - Open archival information system (OAIS) - Reference model* and it is approved by ISO through prior agreement with Consultative Committee for Space Data Systems rather than ISO technical committees.

terms, by extension challenging archives and libraries to address concerns of post-colonialism (Sahlins 1985, Komaroff and Komaroff 1992). It's no small accomplishment to note that for three decades now any number of disenfranchised communities have taken back control of cultural storehouses to establish new and often conflicting histories that subvert established norms and empower those previously excluded (see Choudry and Vally 2017 and references for an overview and introduction). Archives, libraries and museums have largely welcomed these new if at times unruly patrons on the assumption that if the epistemology of the institution is not fundamentally about justice then, by default, its purpose is to sustain injustice (O'Neill 2006).

The tortured history of meaning-making in the late 20th and early 21st century seems strangely at odds with the processes and norms adopted in digital preservation, especially with respect to representation information as described above. If the link to an authoritative definition within a representation network is one of the keys to unlocking meaning, then whoever gets to assign that link or manage the end-point is a very important individual, a dependency that is open to abuse. It's even more problematic when one considers the implied knowledge of the designated community that prevents representation information from spooling out of control. It implies that, if you're not part of the designated community, you're not expected to use or understand the collection, and the library has no explicit responsibility to help you, and no requirement to track your needs. This might be true in the context of academic research such as space science where a relatively small but expert group of professionals would be expected to use complex datasets and would be motivated enough to cope with opaque documentation and annotations. Arguably this is a risky outcome when identities, actions or meanings are in dispute: where honest misunderstanding may arise, or faux conflicts be engineered and prolonged. It is altogether more concerning for public facing institutions whose mission statements articulate noble aspirations for global knowledge through the generations.

Thus, best practice in the digital preservation community means that the digital preservation staff are empowered, in fact required, to exercise a kind of intellectual exclusion that seems out of step with just about every other kind of memory institution.

Libraries have spent 30 years or more coming to terms with inclusion and polysemy: a challenge which the digital preservation community has largely overlooked.

To summarize, good practice in digital preservation seems to be behind the wider library, museum and archive community with respect to community engagement and meaning-making. It will need to adapt and there are signs that this is beginning to happen. In turn the legal deposit regime seems several years behind the digital preservation community in making good on the promise of preservation through purposeful and expansive engagement with users. From this perspective, there seems to be quite a lot of catching up required to ensure that the legal deposit regulations will remain fit for purpose in the longer term.

5. The End and Edges of Print: What is (are) data and what is software?

The non-print legal deposit regime might be described as a first-generation digital process: the regulations model an analogue process with metaphors and workflows concerning journals and books, publishers, libraries and copies. The archive sector has begun to talk about a disruptive process of second-generation digital archives which recognise that collections are digital by default, that workflows and assumptions should be modelled around digital requirements, and that analogue metaphors will need to be set aside if the archive is to remain relevant and coherent (eg The National Archives 2017). This same disruption seems inevitable and perhaps overdue in the context of legal deposit.

The Legal Deposit Libraries Act (2003) and subsequent regulations are quite explicit about what might be collected and what might not: the libraries have the right to request not only an online publication but any computer program necessary in order to access the work as well as any manual that accompanies it. This is not a blanket permission to collect software, but it does recognize that software and data are often inter-twined. But as data, publication and software become more complicated so these assumptions seem increasingly divergent from the practical reality of digital creativity. It implies a false dichotomy between software and data. By extension the lack of a national deposit collection of software whether as published objects in their own right or an integrated component of an archival information package, seems likely to become more problematic over the years.

These rather abstract tensions have been apparent for a while but come to a head with the emergence and widespread adoption of the cloud (for an historical account of cloud computing and its relationship to desktop computing see Hu 2015, for a more accessible description see Naughton 2012).

Libraries typically approach cloud computing as a storage platform (see Hibberd 2017 and references), but storage seems the least important aspect. The cloud is perhaps more important as a utility and service to data creators. The last decade has witnessed a major architectural change between computing as a product and computing as a service which in turn may render obsolete the current provisions of the legal deposit regime.

In a desktop computing environment, files are more or less self-contained on a disk, and canonical versions of software are provided so that a user can create or alter files in a more or less canonical format in which data is wrapped. Files might be remote and may even be spread over multiple file stores with symbolic references and disks: but there is a byte stream that can be assembled and committed as a file which behaves predictably for all users. Likewise, the software is dynamic, with versions and subversions and service packs supporting them: but there is a single stack and all of them are present at a moment in time in a single grey or beige or black box.

In a cloud environment, everything is a service, accessed through a browser, a web connection and a login. Software is summoned on demand, adapting to suit the configuration of device, browser and user, and it will almost certainly have been updated since it was last used. The software and its execution environment are remote and the stack of applications on which the interaction depends is assembled on demand and differently for each user. The same is true of data: services might use the metaphor of a file, but more likely the file is a series of symbolic links to a highly distributed series of byte streams which can be assembled differently depending on user. It may be possible to save this to a local hard disk: but that is just a neat way of synchronising the local environment with distributed remote storage and is in some sense always a migration from the original. There may never be a file in the conventional sense, nor a self-contained programme to run or, or a stable format to be rendered, or a fixed user perspective.

Having spotted that platform, software, infrastructure and even data have become services in the cloud, then it becomes quite easy to think of a remote service being compiled from distributed micro-services. And that is what they are: services behaving as services. There's a sort of fractal inclination implicit here, with micro-services drawing on yet smaller services and so on. It may not carry on *ad infinitum* but can continue as long and as far as the business need exists and supply chain can support. And thus, long chains of interdependence emerge in which multiple components have maintenance cycles that are invisible, at least until they go wrong. By implication all of this will fall into the provisions of article 6.2.b of the Legal Deposit Libraries Act 2003.

For the sake of clarity, an example arises in the context of fonts: is a font a necessary component of a published work? They are a simple accessible example of how the gap between application and data can be hard to sustain, and there are a few examples of what happens when fonts and text are separated by virtue of deliberate acts to restrict their distribution or remote updates that fail (see Kilbride 2017a for three relevant examples in different types of document), underlining a fragility in the current legal deposit regulations. If font and text can be so easily separated, then challenges with relatively well understood formats will certainly be true of complicated ones like geospatial data or multi-user relational databases which also fall within the purview of legal deposit, not to mention complex personalised interactions with web resources, emerging formats for enhanced e-books, or research data supplied in support of e-Journals. If this vulnerability is evident in collections which are accessed every day when people notice unexpected flaws, then it will also be true of preserved collections to which access is restricted and access happens infrequently. If it is true of unexceptional or uncontested data, then it will be even more true of highly contested or high impact systems.

The questions that arise for the preservation of non-print legal deposit collections are profound: what are the limits of dependency; how can preservation services detect such dependencies confidently, and how can they assess or mitigate the risks such dependencies pose? What are the implications for licensing and is there a potential for orphaned micro-services much as there is for orphan works (Korn 2009, Hoeren et al 2013)? How might we

ensure that data is authentic and accessible under reproducible conditions when the underlying services are continuously variable? These are not small questions and they disrupt the print-based and desk-top computing assumptions of the current legal deposit regulations.

6. Digital Preservation, Legal Deposit and the Second Quarter of the Twenty-First Century

This article began by questioning the capability of legal deposit libraries in the UK and Ireland to preserve non-print legal deposit collections, finding that, despite a number of flaws, they have risen to the challenge with energy and skill over the last five years, even while the size and nature of that challenge has changed. Moreover, the culture of collaboration and continuous quality improvement are to be commended. The legal deposit libraries' aspirations to be sector-leading internationally has been independently verified. There is much to celebrate and reason to be optimistic, but there is no time to be complacent: on the contrary the energetic commitment to innovation seems to be single most important characteristic in delivering this success.

The future is less certain. The legal deposit libraries seem well set to fulfil the requirements of the current regulations, though such statements are inevitably dependent on funding and leadership as well as a clear line of communication with publishers. But the same regulations seem contradictory: how can a digital collection be preserved without also having clear lines of communication and significant feedback from the user community; how can the requirement to preserve supporting software be sustained in an environment where software, platform and even infrastructure are offered as services?

There has seldom been a more interesting time for libraries, nor a more challenging one: there seems little prospect of remaining unchanged. In this context, it would be ungenerous not to offer a number of recommendations that will ensure secure access to our common digital legacy:

Non-print legal deposit collections are dynamic, surprising in scale and changing rapidly in configuration. Up till this point the concentration of expertise and dedication to innovation demonstrated at the British Library on behalf of the other libraries has been exemplary: but

this is no excuse to reduce efforts. On the contrary the lesson from the last five years is that only through an ongoing commitment to continuous quality improvement is there any prospect of preserving and expanding non-print legal deposit collections into the second quarter of the twenty-first century.

There is an implicit contradiction between preservation and access embedded within the legal deposit regulations in the UK and Ireland. The legal deposit regime, with severe constraints to access, has inhibited the development of a meaningful audience and as a consequence there is precious little dialogue between preservation actions and designated community. If this is ignored the mission to preserve will be defeated: instead of inspiring the next generation the legal deposit regime could well inadvertently disenfranchise them.

There are unresolved tensions between the public role of libraries to enable communities and support meaning-making which in the context of historic collections are aggravated by the absence of authorial voice. The digital preservation community, concerned with solutions, has been slow to recognise the extent to which it has diverged from the mainstream of cultural theory. The legal deposit regime exacerbates this, inhibiting access, dissociating collections from designated communities and eroding preservation services.

The technological assumptions underpinning legal deposit place unfortunate emphasis on a false dichotomy between data and software. This distinction has never been viable but is rapidly becoming unsustainable in the advent of cloud computing in which software, application and data can no longer be isolated. A paradigm shift in computing means that specific articles of legal deposit face imminent obsolescence.

This article asks, in the face of an uncertain present and more uncertain future, what are the prospects that non-print legal deposit collections of the last five years will be available to future generations. There can be no guarantee. But given the energy and dedication exhibited so far, given the open commitment to innovation, confidence in their mission, and a willingness to address evident shortcomings in the regulations, the legal deposit libraries have created a favourable environment in which preservation is a distinct prospect.

Acknowledgements

I am grateful to Paul Gooding and Melissa Terras for editorial comments and support in preparing this essay and to Paul Wheatley, Sara Thomson and Sarah Middleton for comments on initial drafts.

References

- Adorno, Theodor and Horkheimer, Max 1972 *The Culture Industry as Mass Deception*, in *Dialectic of Enlightenment*, Herder and Herder
- Bantin Philip 2016 *Building trustworthy digital repositories: theory and implementation*, Rowman and Littlefield, Lanham
- Barthes, Roland 1973 *Mythologies*, Granada, St Albans
- Beagrie, Neil 2013 *Preservation, Trust and Continuing Access for E-Journals*, Digital Preservation Coalition Technology Watch Report 13-04 online at: <http://dx.doi.org/10.7207/twr13-04> last accessed 22/10/18
- Brindley, L 2000 Keynote Address to the Preservation 2000 Conference, *The New Review of Academic Librarianship* 6, pp 125-138
- British Library 2010 *2020 Vision*, British Library online at: <http://www.bl.uk/aboutus/stratpolprog/2020vision/2020A3.pdf> last accessed 23/10/18
- Brown, Adrian 2013 *Practical Digital Preservation: a How-to Guide for Organizations or Any Size*, Facet, London
- Center for Preservation and Access / Research Libraries Group 1996 *Preserving Digital Information*, online at: <https://www.clir.org/wp-content/uploads/sites/6/pub63watersgarrett.pdf> last accessed 22/10/18
- CRL, The Center for Research Libraries, OCLC, Online Computer Library Center, 2007, *TRAC, Trustworthy Repositories Audit & Certification: Criteria and Checklist*, available at: <http://www.crl.edu/PDF/trac.pdf> last accessed 11/10/18
- Choudry, Aziz and Vally, Salim 2017 'History's Schools: Past struggles and Present Realities' in Aziz Choudry and Salim Vally (eds) *Reflections on Knowledge, Learning and Social Movements: History's Schools*, Routledge London 1-19

Comaroff Jean and Comaroff John 1992 *Ethnography and the Historical Imagination*, Westview, Boulder

Consultative Committee for Space Data Systems Secretariat 2012, Reference model for an open archival information system (OAIS): Recommended practice (CCSDS 650.0-M-2: Magenta Book), CCSDS, Washington, DC. online at <http://public.ccsds.org/publications/archive/650x0m2.pdf> last accessed 22/10/18

D'Ancona, Matthew 2017 *Post Truth: The New War on Truth and How to Fight Back*, Ebury, London

Day, Michael 2001 'Preservation 2000' in *Ariadne 26*, online at: www.ariadne.ac.uk/issue26/metadata/ last accessed 7/9/2018

Day, Michael, Pennock, Maureen, May, Peter, Davies, Kevin, Whibley, Simon, Halvarsson, Edith, and Kimura Akiko 2016 'The preservation of disk-based content at the British Library: Lessons from the Flashback project', *Alexandria 26* pp 216-234

Day, Michael, Pennock, Maureen, Smith, Caylin, Cooke, Ian and Jenkins Jeremy 2018 Preservation planning for emerging formats at the British Library, Proceedings of the fifteenth International Conference on the Preservation of Digital Materials (iPres) online at: <https://osf.io/65p7m/> last accessed 22/10/18

DCMS 2013 Guidance on the Legal Deposit Libraries (Non-Print Works) Regulations 2013, Department for Culture Media and Sport, online at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/182339/NPLD_Guidance_April_2013.pdf last accessed 18/10/18

Derrida Jacques 1976 *Of Grammatology*, Johns Hopkins University Press, Baltimore

Derrida, Jacques and Prenowitz Eric 1995 'Archive Fever: a Freudian Impression' in *Diacritics* 25, pp 9-63)

Digital Preservation Coalition (2018) Non-Print Legal Deposit Digital Preservation Review, Final Report, online at: <https://www.bl.uk/britishlibrary/~media/bl/global/digital%20preservation/non-print-legal-deposit-digital-preservation-review.pdf> last accessed 11/10/18

Duranti, Lucia 2016 'What will trustworthy systems look like in the Future?' in Philip Bantin (ed) *Building Trustworthy Digital Repositories: Theory and Implementation*, pp 337-8

Gartner, Richard and Lavoie, Brian Preservation 2013 Metadata (2nd edition) Digital Preservation Coalition Technology Watch Report 13-3 online at: <http://dx.doi.org/10.7207/twr13-03> last accessed 22/10/18

Hall, Stuart 1990, *Culture, Media, Language*, Unwin, London

Harvey, Ross and Weatherburn Jaye 2018 *Preserving Digital Materials* (Third Edition), Rowman and Littlefield, London

Hedstrom, Margaret 1998 'Digital preservation: A Time Bomb for Digital Libraries' in *Computers and the Humanities* 31 pp 189-202

HM Government 2003 *The Legal Deposit Libraries Act 2003*, online at <https://www.legislation.gov.uk/ukpga/2003/28/contents> last accessed 18/10/18

HM Government 2013 *The Legal Deposit Libraries (Non-Print Works) Regulations*, Statutory Instrument 2013.777, online at <http://www.legislation.gov.uk/uksi/2013/777/contents/made> last accessed 18/10/2018

Hibberd, Lee 2017 *Cloudy Culture: Preserving digital culture in the cloud* in DPC Blog November 2017 online at: <https://www.dpconline.org/blog/cloudy-culture-part-4> last accessed 22/10/18

Hoeren Thomas, Kolany-Raiser, Barbara and Yankova, Silviya 2013 *Legal Aspect of Digital Preservation*, Edward Elgar, Cheltenham

Hu, Tung-Hui 2015 *A Prehistory of the Cloud*, MIT Press, Boston

ISO (2012) *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories* (ISO 16363:2012 / CCSDS 652.0-R-1) International Standards Organization, Geneva

John, Jeremy Leighton, *Digital Forensics and Preservation*, Digital Preservation Coalition Technology Watch Report 12-03, online at: <http://dx.doi.org/10.7207/twr12-03> last accessed 22/10/18

Kilbride, William 2017a *The Data Vanishes*, in DPC Blog August 2017, online at:

<http://www.dpconline.org/blog/the-data-vanishes> last accessed 12/10/18

Kilbride, William 2017b No End of History in DPC Blog June 2017 online at:

<https://www.dpconline.org/blog/no-end-of-history> last accessed 22/10/18

Kilbride, William 2017c *Obsolescence 2.0: Digital Preservation by People for People*, in *DPC*

Blog 2nd May 2017 online at: <https://www.dpconline.org/blog/obsolescence-2-0-digital-preservation-by-people-for-people> last accessed 11/10/18

Kirchhoff, Amy and Morrissey, Sheila 2014 *Preserving eBooks*, Digital Preservation Coalition

Technology Watch Report 14-01 online at: <http://dx.doi.org/10.7207/twr13-04> last accessed 22/10/18

Korn, N 2009 *In from the Cold: An Assessment of the Scope of Orphan Works and its Impact on the Delivery of Services to the Public*, Strategic Content Alliance, online at:

<https://www.djsresearch.co.uk/Free/published/infromthecoldv1.pdf> last accessed 23/10/18

National Library of Scotland 2015 *The Way Forward: Library Strategy 2015-2020*, online at:

<https://www.nls.uk/media/1190455/2015-strategy.pdf> last accessed 22/10/18

Naughton, John 2012 *From Gutenberg to Zuckerberg: What you really need to know about the Internet*, Quercus, London

O'Neill, Mark 2006 'Essentialism, adaptation and justice: Towards a new epistemology of museums' in *Museum Management and Curatorship* 21, pp 95-116

Pennock, Maureen 2013 *Web-Archiving*, Digital Preservation Coalition Technology Watch

Report 13-01 online at: <http://dx.doi.org/10.7207/twr13-01> last accessed 22/10/18

Pennock, Maureen and Smith, Caylin 2016, *Managing an ISO 16363 Self-Assessment: A How To Guide*, 11th International Digital Curation Conference online at:

http://www.dcc.ac.uk/sites/default/files/documents/IDCC16/18_Managing_ISO16363.pdf last accessed 22/10/18

RLG/OCLC Working group on digital archive attributes, (2002), *Trusted digital repositories.*

Attributes and responsibilities. An RLG-OCLC Report, RLG, Mountain View, CA, available at:

<https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>

last accessed 11/10/18

Rosenthal, D 2014 TRAC Certification of the CLOCKSS Archive, in *DShR Blog* online at:

<https://blog.dshr.org/2014/07/trac-certification-of-clockss-archive.html> last
accessed 8/10/18

Sahlins, Marshall 1985 *Islands of History*, University of Chicago Press, Chicago

Shenton, H 2000 'From Talking to Doing: Digital Preservation at the British Library', in *The
New Review of Academic Librarianship* 6, pp 163-178

Spivak, Gayatri 1990 *The Post-Colonial Critique: Interviews, Dialogues Strategies*, Routledge,
London

The National Archives 2017 *Digital Strategy*, online at:

[http://www.nationalarchives.gov.uk/documents/the-national-archives-digital-
strategy-2017-19.pdf](http://www.nationalarchives.gov.uk/documents/the-national-archives-digital-strategy-2017-19.pdf) last accessed 23/10/18

Webb, C 2000 Towards a Preserved National Collection of Selected Australian Digital
Publications, in *The New Review of Academic Librarianship* 6, pp 179-192