
'An Ark to Save Learning from Deluge'? Reconceptualising Legal Deposit after the Digital Turn

Paul Gooding (University of Glasgow) and Melissa Terras (University of Edinburgh)

Introduction

Despite the introduction of Non-Print Legal Deposit (NPLD), the concept of legal deposit is still viewed primarily as a function of print modes of publishing and consumption. This chapter will argue that the print era notions that influence the NPLD access and reuse regulations are increasingly out of step with broader developments in publishing, information technology, and broader socio-political trends in access to information. We will explore case studies relating to archives of web materials, in order to demonstrate the ways that innovative research, publishing, and copyright are changing our understanding of what constitutes an archive. The digital archive is a space where innovation is occurring, and where the role of the library is evolving, yet NPLD regulations largely close down options for innovative approaches to digital materials.

This chapter will use two case studies which encounter digital materials as a new media form, with their own functions and affordances. First, we will look at the Internet Archive, a non-profit organisation that provides a suite of services to support national library and public-facing web archives. We will discuss how the IA's approach to copyright, and its openness to data-driven methods, allow it to provide web archival services that go far beyond those allowed under legal deposit. Second, we will review Common Crawl, another non-profit organisation that provides web archival content for non-commercial text and data mining. Through these case studies, we will demonstrate that web archives can provide a space for computational research. Innovation with digital materials is occurring within national library labs in particular, and we will conclude by applying the lessons learned from web archives to understand the challenges of innovative reuse of the NPLD eBook and eJournal collections. Similar change is occurring in relation to scholarly publishing, and yet digital scholarship is similarly closed off in each subset of NPLD collections. For scholarly publishing and use, NPLD's remediation of print services leaves it out of step with changing approaches to digital materiality elsewhere. The significance of this chapter is twofold: first, it advocates for a new understanding of access to legal deposit in relation to the textuality of digital media, which we label the 'datafication' of the legal deposit library; and second, it recognises that unlike print media where legal deposit materials were often not the only copy available, many non-print collections will be unique to the legal deposit libraries. We will conclude by suggesting ways in which NPLD regulations could support new approaches to digital materials in a variety of formats, based on aligning legal deposit to UK copyright regulations, but within the confines of the library reading rooms. This process will assist us to understand how to promote innovative reuse of the unique and vulnerable collections preserved by legal deposit, without undermining the commercial viability of publications through unfettered open access.

The Historical Context for Legal Deposit

Sir Thomas Bodley (1545-1613) casts a long shadow over the history of legal deposit in the United Kingdom. His influence began in 1598, when he wrote to the Vice-Chancellor of the University of Oxford with an offer to redevelop the University Library, with the following words:

> Where there hath been heretofore a publike library in Oxford: which you know is apparent by the rome itself remaining, and by your statute records I will take the charge and cost upon me, to reduce it again to its former use (in Philip, 1983, p. 1).

By 1602, the fruits of his work were evident, as the Library reopened under the leadership of Thomas James (c.1573-1629). James was the first Librarian of the Bodleian Library, with the institution renamed in honour of its founder. The significance of Bodley's act of patronage was recognised by his contemporaries. Francis Bacon, for instance, praised him in 1605 for 'having built an ark to save learning from Deluge' (in Spedding, 1861, p. 253).

Legal deposit has existed in various forms, in various nations, for over 500 years. This lineage shows that successive generations have keenly valued the idea that national memory can be protected through the capture and preservation of our published history. With the introduction of NPLD in the UK, and similar regulations elsewhere, it is evident that the value of regulations for formal deposit remains broadly recognised by government, librarians, publishers and researchers alike. These stakeholders broadly recognise the broad prestige and posterity value of NPLD, which associate it with the same intrinsic value as print legal deposit collections (Gooding, Terras and Berube, 2019, p. 17). Iain Sproat MP, for instance, noted in a 1997 parliamentary debate on extending legal deposit to non-print materials that 'the main purpose of legal deposit is to establish as comprehensive an archive as possible of our national published output for use by future generations' (HC Deb, 1997). Yet this is not its only purpose.

Lariviére (2000), for example, considers legal deposit to be the foundation of a national policy to support freedom of expression and access to information, while Brazier points out that the resultant collections form the basis for the world's 'great research libraries' (2016, p. 42). De Beer et al. further outline several ways in which the creation of comprehensive legal deposit collections can support the national interest:

> It is important to provide citizens as well as researchers (within the country as well as abroad) with access to a research collection of the country's published material, it supports bibliographic control and it makes it possible to monitor the growth of the publishing industry (De Beer *et al.*, 2016, p. 88).

In many respects, then, Bacon's 'Ark to save learning' appears as deft a metaphor for contemporary legal deposit as it was for Bodley's acts of patronage over 400 years before. Bacon's words hint at the public service remit of the regulations, the effort to capture at least some of the modern data deluge, and the focus on long-term preservation through which we understand legal deposit. On the other hand, the contemporary metaphor of the 'digital universe' (National Library of Scotland, 2013) that greeted the introduction of NPLD in the UK recalls those utopian visions of the universal library that have permeated our thinking all the way from the Renaissance through to the mission of the Internet Archive (Kahle, 2007). Gantz and Reinsel describe the true breadth of the so-called universe:

> Images and videos on mobile phones uploaded to YouTube, digital movies populating the pixels of our high-definition TVs, banking data swiped in an ATM, security footage at airports and major events such as the Olympic Games, subatomic collisions recorded by the Large Hadron Collider at CERN, transponders recording highway tolls, voice calls zipping through digital phone lines, and texting as a widespread means of communication (Gantz and Reinsel, 2012).

Legal deposit, by comparison, preserves a razor-narrow subsection of this flood of digital information: eBooks, eJournals, electronic mapping, CD-ROMS, newspapers and websites are

all included, whereas the data deluge described by Gantz and Reinsel contains nothing that falls under legal deposit. Whether these data are *worth* preserving is a different matter, but for now it is enough to be clear that they are *not* preserved, at least not systematically for the public good. For this reason, we intend to focus solely on digital forms that come under NPLD, in order to explore how concepts of digital materiality inform our thinking around the scope of the regulations.

It is foundational to this line of thinking to understand that digital materiality offers different affordances than those offered by printed media. Indeed, N. Katherine Hayles argues that:

> Our notions of textuality are shot through with assumptions specific to print, although they have not been generally recognized as such. The advent of electronic textuality presents us with an unparalleled opportunity to reformulate fundamental ideas about texts (Hayles, 2005, p. 89).

One way in which she understands this is through the 'work as assemblage' (Hayles, 2005, pp. 104–109), a notion that echoes Lev Manovich's (2001) description of the modularity of digital media. The automation of tasks, and the ability to apply computational methods to digital data, are based upon the ability of digital texts to be described formally, and to be understood as assemblages of their constituent elements. Hayles argues that the context and presentation of a text is key to its meaning, and that there is therefore a spectrum of similarity and difference, with clusters of similar textual embodiments of a specific work emerging. Texts can therefore be discussed both in terms of their content and their physical embodiment. This approach to digital textuality supports the idea that archived websites are formally unique from the original website because they draw on modular components to create an interpretation rather than a facsimile copy:

> The archived website is a reconstruction in the sense that it has to be assembled by the use of all the archived bits and pieces, first when they are archived, and later when the material has to

> be displayed for the user of the archive. Thus, it could be argued that the archived website did not exist before it entered the archive, and in this respect it differs significantly from other media types (Brügger, 2012).

Shifting forms of digital textuality have implications for historians engaged in researching the late twentieth and early twenty-first century, an era for which digital media are primary resources. Winters notes that search and discovery tools are not adequate for the kinds of analysis that are required of web archival sources, thereby distancing research from print-era discovery and reading models:

> The dominance of (a particular type of) search as a digital research method quickly becomes problematic for web archives where, quite apart from difficulties arising from scale, the scope of a particular archive is unknown and the process of creation largely undocumented. Discovering what might be in the archive is often the primary objective – and this is not well served by keyword searching which produces a list of results unordered by anything other than date (Winters, 2017, p. 243).

We can thus begin to see an emerging understanding of digital textuality, based on the separability of content or data, and the physical embodiment of the text. These multiple layers of meaning require different methods, analytical frameworks, and discovery tools than print resources. One defining feature of digital media is that they can be understood as both data *and* text. As a result, scholars have increasingly begun to engage with digital scholarship, excavating meaning through analysis, synthesis, and visualisation.

The library sector was slow to move beyond an understanding of digital materials as mere carriers of information, arguing that content matters more than format (e.g. Quint, 2001) while continuing to understand new media forms through the language of print. The same logic has been applied to the NPLD regulations, which embody a belief that access to digital materials should mirror the affordances offered by print. Such thinking ignores material shifts,

other than where preservation of content might be affected: Seadle, for instance, asserts that 'for digital materials, it makes no sense to write rules for legal deposit based on the medium. Increasingly the medium on which a digital work exists matters less than what mark-up format it uses, what external links it requires, and what technological protection it has' (2001, p. 302). He is correct, but only insofar as it would be impractical to develop regulations that respond directly to individual digital forms. However, since Quint and Seadle asserted the primacy of content over form in 2001, there has been a reassertion of the importance of textuality in the digital age. The assumptions of print are explicitly incorporated into NPLD regulations, although they are not always recognised as such.

The doors to the Ark: Access to UK NPLD materials

It is necessary for legal deposit regulations to define boundaries that make it feasible for libraries to undertake collection and preservation activities, given their limited funding and often small operational staffing. However, the print-based assumptions of legal deposit in the UK are important precisely because they continue to define NPLD; indeed, the broad scope of the digital universe stands in stark contrast to the access model that applies to NPLD materials[i]. The NPLD regulations were introduced in 2013 to address a twofold challenge: a decrease in the number of print publications being collected under legal deposit; and a dramatic increase in the amount of born-digital content that, due to its publishing format, would no longer fall under legal deposit regulations. Print intake under legal deposit has steadily fallen, and indeed given the huge success of NPLD in collecting other materials it can be easily surmised that print materials now form an ever-smaller subset of published materials. As long ago as 1998, the working party on non-print legal deposit concluded that it was necessary to extend statutory deposit in order to secure a comprehensive national collection (British Library, 1998). The working group paved the way for an interim voluntary arrangement, which was introduced in

2003 and allowed the UK deposit libraries to make significant progress in archiving non-print materials. This voluntary arrangement was underpinned by the 2003 Legal Deposit Libraries Act, (2003) which established the principle of electronic legal deposit and provided legal protection to the deposit libraries to allow them to collect electronic materials at scale; however, it also required a further piece of subordinate legislation to enact its provisions fully in law. In order to establish a consensus, the Legal Deposit Advisory Panel was convened from 2005 to 2010. In 2009, LDAP submitted recommendations for the legal deposit of offline media, and the harvesting and archiving of web materials. In 2010, it made further recommendations, covering paywalled electronic materials, structured datasets such as railway timetables, and content which is pushed to the user by electronic means (Gibby and Brazier, 2012). After a period of consultation, the introduction of The Legal Deposit Libraries (Non-Print Works) Regulations 2013 (2013) formalised arrangements for collection, preservation and providing access to NPLD materials.

These regulations were the result of many years of careful negotiation and thought, and have been a huge benefit to the legal deposit libraries over the last five years, but it is worth exploring how the assumptions that underpinned their introduction may be problematic when we view deposited works as data (Padilla *et al.,* 2019). The current interpretation of NPLD exists as a print-era regulation adapted for a prescribed range of digital materials, without mechanisms to deal with changing material, formal and structural considerations. Nowhere is this more evident than in the formulation of access in the regulations. There are two layers to the access criteria: first, the broad wording of the regulations that define access and use; and second, the way that legal deposit libraries have interpreted and implemented those stipulations. In this respect, little has changed since Bodley's time, with the reference-only policy of the Bodleian Library seeing just a few users each day using its resources (Bodleian Libraries, 2017). Access is still defined in respect to the single copy deposited under print legal

deposit, which due to its format could only ever be accessed by one individual at a time. Indeed, when Lord Gardiner of Kimble reported to the House of Lords in 2013, he explained that reader access to NPLD materials had been regulated 'in order to mirror the level of access to printed publications' (HL Deb, 2013). The regulations (2013) state strict rules for how materials may be accessed:

- Reader access to NPLD materials is limited to computer terminals on premises controlled by the legal deposit libraries (Part 1, Regulation 2).
- Each legal deposit library must ensure that material is only accessible to readers via one computer at a time (Part 4, Regulation 23).
- For materials published online, at least seven days must elapse between the date of delivery of that material, and the date on which it is made available to readers (Part 4, Regulation 24).
- A copyright owner may request in writing that access is withheld for a specific period of time. Deposit libraries are bound to comply with such requests, provided that:
    - The period for which materials are withheld does not exceed three years from the date on which the request is made;
    - The deposit library is satisfied that, for the specified period, viewing of the relevant material by readers would, or would be likely to, 'unreasonably prejudice the interests of the person making the request' (Part 4, Regulation 25).

In some respects, the deposit libraries adopt a consistent interpretation of the regulations. The agreed technical solution for preserving NPLD materials is based on a 'Shared Technical Infrastructure' (British Library, 2013), housed in a secure environment with no public access[ii]. Each of the national libraries stores a full copy of all NPLD materials at a local node, based in St. Pancras, Boston Spa, Aberystwyth and Edinburgh, while the academic legal deposit libraries each connect to the British Library nodes. The system is designed to be secure, to

protect against unauthorised use, and to be resilient enough to ensure long-term preservation of assets. The nodes are also responsible for managing access to ensure only one copy of each unique item is used concurrently. All libraries share a technical solution called ERICOM that delivers digital materials through a 'secured remote desktop browser system' which ensures that files are not stored locally: it is to all intents and purposes streamed remotely, to protect against the proliferation of illegal copies (British Library, 2013). This is a solution agreed on by publishers and libraries, as appropriate to ensure the security of deposited materials.

The UK's legal deposit libraries have each been responsible for working out how these regulations would be implemented locally, and so the extent of access for users is further defined by the location of computer terminals. The British Library provides fixed terminals only within its reading rooms, for instance, as does the Cambridge University Library, limiting access to a specific subset of fixed terminals within each library.[iii] The Bodleian Libraries, on the other hand, allow access via any fixed terminal, as long as the IP address is recognised to reside within the library's walls. The National Library of Scotland has gone further still, creating a digital reading room in Glasgow's Kelvin Hall that expands fixed terminal access to its digital collections beyond its historic home in Edinburgh (National Library of Scotland, 2016).

The NLS decision to allow users outside Edinburgh to access electronic materials closer to home, emphasises how competing conceptual frameworks can come to influence the scope and effectiveness of the 2013 regulations. It is a matter of public record that models for access to NPLD materials are shaped largely upon the level of access to print publications. Here, we make the less obvious point that those access models are built on assumptions that are specific to the print medium but have been transplanted over centuries to become foundational tenets of legal deposit. These tenets, now in the process of being reformulated, are largely responsible for ensuring that the textuality of digital media is ignored in a legal deposit environment.

NPLD regulations, and the interpretation of them at local level, artificially mimic the inherent scarcity of the printed form. One of the key tenets of legal deposit is that access to NPLD materials must protect the commercial interests of rights holders, and while this is likely to remain integral to any future regulations it is the problems that arise when defining reference access in terms that relate specifically to printed texts. Print works most conveniently lend themselves to traditional models of scholarship developed around information discovery via catalogue searching, followed by reading or browsing of specific volumes. These traditional models grew up alongside print, in order to make the most of the affordances of the available materials; however, digital materials have become increasingly understood in terms of their own textuality. As we have argued in the past (Gooding, 2016), it is important to recognise that digital materials are different: they provide a different experience to users that is nevertheless rich when understood on its own terms. Digital materials allow us to make new links between resources, to search huge amounts of material with a few key presses, to develop new forms of scholarship, and to make derivative and public domain works available online. These are all vital functions of digital materiality that are unavailable to users of NPLD materials, who have increasing expectations of open access to data, remote access to library resources , and the ability to perform text and data mining (Winters, 2017, p. 46; Gooding, Terras and Berube, 2019, pp. 18–23).

New Media, New Users, New Services

This is not to say that libraries have been totally unable to provide innovative services to their users. In recent years, libraries have been experimenting with services that have allowed them to respond to the needs of users of digital materials. Like many other national libraries, the British Library has responded by developing a digital scholarship team with direct responsibility for collection areas with a strong digital practice focus, internal digital skills

training and dissemination, and knowledge exchange with researchers. This has led to training programmes in digital research methods (McGregor and Farquhar, 2013), new staff roles, and particularly the emergence of library 'Labs' initiatives (Gooding, 2017, pp. 100–101). Brooks et al. define library labs as 'any library program, physical or digital (or a hybrid) in which innovative approaches to library services, tools, or materials are tested in some structure way before being made part of regular workflows, programs, or mission' (2013, p. 186). Labs emerged from a culture of experimentation with supporting digital research methods. For instance, the British Library Dataset Programme saw the creation of DataCite, the international data citation initiative which assigns Digital Object Identifiers to datasets. These services were then leveraged to promote dataset discovery services (Wilkinson, Pollard and Farquhar, 2010). Other organisations, such as the National Library of Wales, also developed digital research teams to provide core expertise and knowledge exchange, and to undertake in-house research. These activities demonstrate a willingness among libraries to engage with their digital collections, to make them available to users, and to provide innovative services that facilitate computational research with library collections (Gooding, 2017, p. 103).

By providing a defined space, and often dedicated staff with specific expertise in programming, digital collections, and digital research, library labs can help foster a culture of innovation and experimentation. For instance, the British Library (Kremerskothen, 2013) and the National Library of Wales (Pugh, 2010) have experimented with releasing digitised images to the commons via Flickr, to great success (Scholz and Miles, 2015), while the National Library of Scotland recently launched its Digital Foundry as an online repository of data for researchers (National Library of Scotland, 2019). The success of labs, usually on limited project funding, has led to their widespread adoption and the emergence of an international library labs network via workshops hosted by BL Labs (Mahey, 2018). It is therefore certainly the case that libraries are developing creative, innovative and inspiring new ways of working

with their collections. However, these activities must generally exclude legal deposit materials, due to restrictions on their use. When it comes to legal deposit, then, we must look towards non-governmental organisations for examples of services that are emerging around new forms of digital media such as web archives.

The following case studies will address two key risks to stakeholders as a result of the framing of access within NPLD: first, to libraries, who risk being left behind by non-governmental innovators in the sector because of the lopsided nature of limitations on access and reuse; and second, to publishers who risk pushing users away from libraries and towards less tightly regulated services in an otherwise understandable attempt to protect their commercial interests. Larsen notes that 'in the paper world legal deposit and preservation of printed heritage are almost synonymous with libraries. In the digital world it is not a matter of course that libraries are best suited to perform these tasks' (2005, p. 86). This is evident in the way that non-profit organisations now fulfil traditional library roles without being subject to legal deposit regulations or the need to adopt risk averse approaches in order to maintain positive publisher-library relationships. As a result, they can bypass certain protections entirely by exploring the boundaries of copyright and IP regulations and experimenting with key exceptions. The high-profile Internet Archive, for instance, uses exceptions to copyright law to allow users far more liberal access to archived web materials than is possible via legal deposit. The less famous Common Crawl leverages the particular strengths of digital textuality in providing a representative sample of global web archival data for text and data mining.

We will conclude by proposing that NPLD regulations must find a balance between protecting publishers' rights and adapting to contemporary innovations in data science and digital scholarship. While some have argued that NPLD should avoid a media-centric approach, on the basis that the material carrier of information is increasingly irrelevant (Seadle, 2001), we will argue that an awareness of the textuality of digital sources is in fact vital for

how they are interpreted, accessed and reused. The NPLD regulations are not media-agnostic: on the contrary, they are fundamentally rooted in structures that deliberately remediate print codes as closely as possible; by artificially recreating the scarcity of the printed form, but also by recreating the inherent limitations to reading that are imposed by print. This important reconceptualisation of legal deposit, which we call the 'datafication of the legal deposit library,' will assist in preserving the precious relationship between publishers, libraries and researchers that has evolved around our national memory over the last five centuries. As the following case studies will show, innovation of this type is already happening outwith the legal deposit libraries.

The Internet Archive: Access, Copyright, and the Limits of Legal Deposit

The Internet Archive is a non-profit organisation based in San Francisco, USA. Founded in 1996, its institutional mission is to 'provide Universal Access to All Knowledge,' an ambitious agenda described by Brewster Kahle as one of the greatest possible contributions to humanity:

> Universal access to all knowledge is possible, and I'd say it could be measured as one of the great achievements of humankind, along with putting a man on the moon or assembling the Library of Alexandria. I think our generation could bring universal access to all knowledge, and that's something we'd be proud of for centuries (Kahle, 2007, p. 31).

The institutional mission of the IA is hugely ambitious, but it is more specific in positioning itself as a 'provider of web archiving technologies and services' (Hockx-Yu, 2016, p. 3), incorporating:

- Open source software for crawling and public access;
- A global web archiving service for the general public;

- Archive-It, a subscription service for creating, managing, accessing and storing web archive collections, and
- A tailored broad crawling service for national libraries (Hockx-Yu, 2016, p. 3).

The UK Legal Deposit Web Archive (UKLDWA) is supported by IA software, with the organisation's Wayback Machine providing the basis for its search interface; however, the IA also makes the Wayback Machine service freely available online to the general public. It has been hugely successful, attracting over 600,000 visitors per day in 2016 (Hockx-Yu, 2016), compared to an average of just 226 visitors per month to the UK Legal Deposit Web Archive (Gooding, Terras and Berube, 2018). Additionally, the IA offers various Application Programming Interfaces (APIs) that allow users to access and build additional services that are able to utilise the IA dataset. Its collections are broader than national libraries, being global rather than national in scope, leading Meyer et al. to remind us that 'while the IA is the *most* comprehensive archive available of web materials, this should not be confused with thinking that the IA crawls represent a *fully* comprehensive record of the web' (2017, p. 27). Unlike the UKLDWA, the IA is also freely available online. For these reasons, and because the UKLDWA is relatively inaccessible by comparison, the IA has become the go-to source for academic researchers interested in accessing historical web content. Brügger and Schroeder's edited volume, for instance, contains regular citations of the IA as a data source or service provider for academic research (Hale, Blank and Alexander, 2017, pp. 45–61; Meyer, *et al.*, 2017, p. 28; Weber, 2017, pp. 83–100), whereas the UKLDWA is presented solely as an object for critique (Meyer, *et al.*, 2017; Winters, 2017). The reliance of UK-based researchers on the IA is evident beyond the confines of this single volume, with the notable exception of studies based on the JISC UK Web Domain dataset. As such the IA, not the UKLDWA, is the *de facto* research service for scholars of the UK web. This is based on the following two key points of difference between the IA and national library legal deposit services.

1.) The Internet Archive provides free access to its collections via the web.

Unlike legal deposit libraries, which are subject to maintaining restrictive access conditions, the IA undertakes its work under the purview of copyright law, as an independent non-governmental organisation. This allows them to take a more liberal approach to copyright, as can be seen in the case of the 'Sonny Bono Memorial Collection,' which was digitised by Elizabeth Townsend Gard, a copyright scholar at Tulane University. Townsend Gard explains that the so-called Last Twenty exception, Section 108(h) of the US Copyright Act, allows published works in the US to be digitised and distributed by libraries, archives and museums, as long as there is no continuing commercial sale of the work, nor availability of reasonably priced copies (Townsend Gard, 2017). With the help of her students, Elizabeth Townsend Gard utilised the exception to digitise a collection of eligible out-of-print books , naming the subsequent collection after 'the author of the bill making this necessary'[iv] (Kahle, 2017). The existence of the collection shows that the IA is willing to explore the boundaries of copyright law to enable access to collections in its care. While there are problems in viewing global archival collections solely through the lens of US copyright law, it is certainly true that the resulting IA services are enabling new forms of research on a global basis. The alternative, as Martin Eve (2016) points out, is that researchers interested in text and data mining in the UK are often forced to either illegally break Digital Rights Management software, or digitise texts themselves. The systems that underpin access to NPLD make this impossible.

2.) The Internet Archive makes it possible for researchers to build services that allow for large-scale computational analysis of web resources.

As a result of its lack of restrictions, The IA has been able to experiment with allowing researchers to undertake computational analysis on web archive resources. Ian Milligan, for

instance, points to the ‘Wide Web Scrape’ (2016), a 2012 experiment by the IA to collect around 80Tb of

files containing around 2.7 billion URIs. The resultant dataset is not freely available online, but is managed through direct queries to the IA in the spirit of supporting non-commercial research (Rossi, 2012). Libraries have partly been left behind for technological reasons, although the growth of Library Labs outlined above points to a way forward. Coyle, for instance, makes it clear that innovation in organisation technology has often occurred within libraries in the past. However, she argues that several factors have had a negative impact on relative levels of innovation in library and information services in comparison to other industries:

> Looking at the relative timelines, those of the overall information technology space and that of the library technology space, it becomes clear that libraries have failed to make the same changes that were happening in the other communities making use of computing. The reasons behind this are undoubtedly many, from issues of budget limitations, institutional conservatism, and the historicity of the library mission (Coyle, 2017).

The innovation gap is partly forced upon libraries by the retrograde assumptions that run through the legal deposit regulations. This is a risk when we bear in mind that web users are often agnostic about their sources, using those that will support retrieval of information or data with as little friction as possible (Warwick *et al.*, 2008; Tanner, 2013), but despite this government and the scholarly community still entrust national collections to the library and archival sector due to their trusted role in society, and long-term view of custodianship. This is important because it distinguishes the ongoing value of the legal deposit libraries as trusted, broad and deep repositories of our printed national record. However, libraries are being left behind in comparison to key newcomers in the library space in terms of access, usability, and provision of tools and datasets. The IA’s Wayback Machine is more accessible and

comprehensive than any publicly available web archive, while users of the UKLDWA face major barriers in accessing NPLD materials for data-driven purposes. As a result, libraries risk the erosion of this trusted position because other providers can respond proactively to changing user needs, take into account regulatory shifts, and take an organisational view on risk.

Common Crawl: The Changing Needs of Researchers

Common Crawl is more explicitly targeted to the needs of data scientists and computational researchers than the Internet Archive. Its approach to web materials raises important questions about our concepts of digital textuality, and how this affects the needs of researchers. Common Crawl was founded in the USA as a non-profit organisation in 2011 with the mission to crawl the web and provide a representative sample of web domains at no cost for non-commercial analysis and research. The organisation undertakes a monthly web crawl which it makes freely available to download as WARC files.[v] Its compliance with copyright regulations is predicated on the US concept of fair use: it argues that websites are intended for human consumption one at a time, whereas Common Crawl undertakes a transformative process by bundling billions of pages together into specialised formats that include text, metadata, and raw data. Furthermore, it explicitly aims to provide a representative sample of roughly 3 billion web pages rather than a comprehensive archive of the global web; in this respect, its mission is very different to that of either the Internet Archive or the national libraries.

The entire data collection model for Common Crawl emphasises that digital materials, and specifically webpages, are data files underneath their representational form. In separating data from form, the organisation is better able to meet the needs of researchers. Winters, for instance, emphasises that ‘it is the portability of data, its separability from an easy-to-use but necessarily limiting interface, which underpins much of the most exciting work in the digital humanities’ (2017, p. 246). Fields that adopt digital approaches to materials rely upon open

access to portable datasets that will allow analysis via computational tools. That legal deposit materials are neither easy-to-use, nor portable, means that libraries are unable to support these researchers in either the short or the long term. As a result, it is easy for libraries to be presented as unwilling to update their practices. Sara Crouse, Director of Common Crawl, does exactly this when she draws attention to:

> The risk averse nature of the web archiving community as a whole (historically many adhered to and still adhere to a strict "opt-in" policy requiring prior approval before crawling a site) and the unwillingness of many archives to modernize their thinking on copyright and to engage more closely with their legal community in ways that could help them expand fair use horizons (in Leetaru, 2017).

While there is an element of truth to this, Crouse misses the extent to which the hands of librarians and archivists are tied by regulations that do not apply to non-governmental content holders. National libraries have released public domain material to the Commons when possible (Pugh, 2010; Kremerskothen, 2013), and they have experimented with open licensing to support research and creative reuse. The independence of Common Crawl and IA is exactly what allows them to innovate, removing them from the problematic print paradigm at the heart of NPLD. However, some NPLD materials are unique, unlike most print legal deposit items, and there will come a time that some digital materials are only available through the legal deposit libraries. It is at this point, in the short to medium term, that an impoverished position on digital textuality could become a barrier to research.

Web materials have been collected under formal legal deposit for just five years but provide a clear example of the differing textuality of digital media. Users are unable to access legal deposit collections in the way that they expect, and in this respect, it could be argued that NPLD is working as intended. It is right and proper that libraries protect the interest of publishers and other rightsholders, and the current regulations are designed to reassure

publishers that legal deposit will not undermine their legitimate interests. In this sense, user expectations are less relevant than the fact that users are denied the opportunity to exercise their legitimate rights under copyright law, in relation to legal deposit materials. This is starkly evident for web archival materials, where it is effectively impossible for users to treat the UKLDWA as a dataset. However, the implications of this print paradigm are felt elsewhere. In particular, even the simple task of accessing Open Access materials is complicated by the framing of NPLD.

Open Access and Copyright

The NPLD regulations aim to mirror levels of access to print legal deposit collections, to reassure publishers that their legitimate interests will not be undermined. However, the digital turn has precipitated not only a reassessment of how we view textuality, but a shift in attitudes towards access that are keenly visible in the Open Access movement in academic publishing. As Adrienne Muir notes elsewhere the importance of Open Science, and Open Access, has been ascribed increased importance in government policy and the scholarly community in the past decade. The UK Research Councils have had policies on Open Access since 2005, and the RCUK Open Access policy of 2013 recognises that free and open access to publicly funded research is a societal good (UK Research and Innovation, 2013). The Open Access requirement adopted for the 2021 Research Excellence Framework formalised an expectation that researchers in Higher Education should increasingly publish in the most open formats possible (UK Research and Innovation, 2013). Most recently, 11 national research funding organisations, with the support of the European Commission and the European Research Council, announced the launch of cOAlition S,[vi] built around the ambitious Plan S, which was described as follows:

> By 2020 scientific publications that result from research funded by public grants provided by participating national and European research councils and funding bodies, must be published in compliant Open Access Journals or on compliant Open Access Platforms (Science Europe, 2018).

The key principles of compliance with cOAlition S include: authors retaining their copyright; incentives for establishing new journals and platforms; institution and funder support for OA fees, alongside a capping of those fees; and a longer-term aim to make monographs and books available via Open Access (Science Europe, 2018). These developments map the start of a trajectory towards a culture of Open Access by default for academic research.

Shifts in UK copyright law further reinforce the exceptional nature of academic research in the regulatory environment, as they embed specific exceptions for non-commercial research. UK copyright law already contains explicit exceptions that allow limited reuse of some copyrighted materials for non-commercial research, or genuine private study. This reuse falls under the UK concept of 'fair dealing,' although the Intellectual Property Office makes it clear that 'there is no statutory definition of fair dealing – it will always be a matter of fact, degree and impression in each case' (Intellectual Property Office, 2014). Tests for legitimate fair dealing include whether the market for the original work would be affected by a proposed usage, and whether the amount of copyrighted work to be taken would be considered reasonable and appropriate. In 2015, the government introduced a further exception that allows researchers to make copies, without specific permission, of whole copyright works to which they have access, for the purpose of computational analysis for non-commercial research (Intellectual Property Office, 2014). This exception does not apply to users of NPLD materials (Gooding, Terras and Berube, 2019, pp. 23–24).

As a result of the digital turn, not only are users increasingly interested in the data that sits underneath our digital materials, but Open Access to publications and data is increasingly

common. Open Access journals commonly use Creative Commons licenses that explicitly allow others to legally build upon, share, and create derivative works under explicit but generous conditions (Eve, 2014, p. 12). For material published under these licenses, publishers will no longer hold rights in the same way as in the past, especially in situations where copyright is not transferred from the author. Furthermore, a publishing model has emerged that is reliant on upfront payment of Article or Book Processing Charges to cover costs and provide profits. These APCs and BPCs often cost up to a few thousand pounds in high impact journals, and ten thousand pounds or more for OA books. In this respect, publishers have no legitimate rights as owners of openly licensed content: indeed, in cases where scholars retain their own copyright, there is a convincing argument that restrictive legal deposit of OA materials in fact infringes upon the legitimate rights of authors to ensure the openness of their publications (Gooding, Terras and Berube, 2019).

The shift towards ubiquitous Open Access in academic publishing has important implications for NPLD, which continues to limit access based on publishing paradigms that are increasingly outdated. Indeed, the phrasing of the regulations introduced a form of 'perpetual copyright' (Gibby and Brazier, 2012) because the legal deposit access restrictions remain in force forever, including after the expiry of copyright. Not only are legal deposit materials not accessible to new forms of research while under copyright, but they will effectively never enter the public domain. This perpetual copyright is a function of the 2003 Legal Deposit Act, which by default excludes all uses of deposit materials unless explicitly permitted by legal deposit regulations (2003, Section 7). This in effect delinks legal deposit from copyright and IP legislation, so that to vary the terms of access it would be necessary to vary the legal deposit regulations. As a result, the restrictions become indefinite, other than by variation through explicit act of parliament. In the scholarly space, at least, the problem of perpetual copyright allows publishers to demand permanent control over NPLD materials even in cases where they

hold no legitimate interests. Users are unable to access legal deposit materials in several important ways that are otherwise enabled by the digital turn: via remote access arrangements; for the purposes of text and data mining; or under Open Access licences. The effect is twofold; first, libraries are unable to serve their contemporary users in accessing digital materials in ways which take advantage of the digital textuality of NPLD materials; and second, users are pushed towards non-governmental organisations which are already providing the precise kinds of data collection and access from which NPLD is supposed to protect the publishing industry.

The different textual and formal affordances of digital texts are contributory factors in a cultural shift towards open access, and towards an understanding of digital media not only as a formal representation of a text, but also as computational data (Padilla *et al.*, 2019). These changes have realigned how researchers and government view access to digital materials, and national policy has recognised the significance of Open Access and computational methods via a series of exceptions that enable their use across the majority of publications. User expectations are thus being shaped by services such as the IA and Common Crawl. They increasingly want information in portable datasets and on demand. In the United Kingdom, though, NPLD restrains users from accessing materials via legal deposit collections, and libraries from providing innovative responses to emerging trends in research and user behaviour, while non-governmental organisations are free to engage in these activities without the responsibility of acting as trusted national repositories. The conceptual fixity of legal deposit, which explicitly informs non-print regulations, does not allow these trusted repositories to evolve in line with broader social, political, and technical trends.

Discussion

We have drawn attention to areas where NPLD has already disadvantaged libraries and contributed to publishers losing control of the data they are trying to control. It is certainly true

that the problem caused by perpetual copyright has been recognised[vii] but there seems little urgency to address it in the short term. In fact, the response to concerns about the inadequacy of NPLD for contemporary users has been to emphasise the posterity-driven mission of legal deposit. Lord Gardiner, for instance, addressed the issue in the House of Lords:

> The Government recognise that the scenario of restrictions on access to content following the expiration of copyright is a concern for the research community. This is an important issue, but will only arise once the copyright term of 70 years has ceased, so in practice the issue will not affect legal deposit for many years to come (HL Deb, 2013).

However, perpetual copyright is just one problem that affects legal deposit; the challenge of Open Access, and the growth in data-driven research, both call into question the print paradigm that underpins the regulations. The result is not just that legal deposit regulations begin to seem anachronistic, but that the protectionist ethos is actively undermined by broader regulatory shifts. Publications that NPLD is tasked with protecting, such as web pages, books and journals, are already available elsewhere, legally, without the continuity that is assumed of national libraries. The services that legal deposit libraries cannot provide for web archives are already being provided by other organisations that operate outside existing legal deposit frameworks.

As a result, the UKLDWA is largely ignored by researchers in favour of resources such as the Internet Archive and Common Crawl. Menell makes a strong case that public access to accurate digital records is an important public good:

> Democracy-enhancing spillovers are particularly important in the digital age. While the rise of the Internet has opened up communications channels to a much greater diversity of speakers, it has, at times, also produced a polarizing cacophony. Expanding access to the most authoritative sources of information and enabling much more accurate and efficient search capability holds the potential to improve the quality of information available. Better information has the power to sharpen and clarify discourse (Menell, 2007, pp. 1042–1043).

The reference-only core of NPLD is an important factor in maintaining publisher support for exactly such authoritative collections, but the great challenge for legal deposit is that the logic at the core of this negotiated compromise has been eroded by changes in copyright, academic publishing, data science, and digital research. Because of this, the challenges to legal deposit cannot simply be passed onto future generations to solve but must instead necessitate a rethinking of the way that reference access is provided. We argue here that the solution is for the 'datafication of the legal deposit library,' which is built on the understanding that key legal deposit reference collections including the UKLDWA have been impacted by the digital turn in ways that undermine the assumptions that underpin their protections. This datafication is evident in the Labs model that many national and research libraries have adopted. The labs allow users to engage with library collections not only as information, or as artefacts, but as data. To address this need, two key areas of legal deposit require addressing:

1.) Alignment with copyright.

Legal deposit libraries face huge challenges because legal deposit regulations are decoupled from copyright. This has created the problem of perpetual copyright and has left libraries behind non-governmental organisations in their ability to provide innovative services to their users. We propose that legal deposit could be more closely aligned to copyright, by specifying exemptions in legal deposit based on access location, rather than access model. In other words, libraries could, if they wish, develop 'Legal Deposit Labs' which allow reuse of legal deposit materials for non-commercial purposes within the physical confines of their reading rooms, or in trusted data safe havens which provide access to computational infrastructure. This would allow libraries to respond proactively to future developments, applied through the prism of reference-only services within libraries.

2.) Open Access and changing IP rights.

As the trend towards OA publishing continues to grow, the idea that legal deposit limitations are in place to protect rights holders' legitimate interests becomes increasingly difficult to defend in the academic publishing space. However, Seadle's (2001) point that the regulations should not be media-centric but general and adaptable requires this to be addressed at the rights-level rather than content level. Exceptions that allow legal deposit libraries to provide access to genuine OA materials would allow legitimate publisher interests to be defended, while ensuring that openly licensed materials are made available in line with rights holders' wishes. It is particularly important to differentiate between material that is Open Access, and that which is only freely available; a particular challenge for the UKLDWA where advertising models often support free access to copyrighted materials. We are also aware that this would place a burden on libraries to create and police differentiated access and reuse systems: this issue would need to be addressed through the implementation stages but would truly allow us to enter an era where legal deposit reference materials are viewed as both content and data.

Conclusion

We have argued in this chapter that the current conception of legal deposit in the United Kingdom relies upon a framework explicitly drawn up to mirror the norms of print legal deposit collections. These norms remediate the textuality of print, with its inherent scarcity, and its suitability to particular forms of information behaviour, defining access to NPLD materials. Libraries and librarians have tended to downplay the significance of digital materiality in defining usage, while often being blind to the fact that print norms can define library services to the detriment of other forms. It is the dual nature of digital media – as content, and as data – that requires us to understand that print is not a neutral form. First, the digital turn has led to a shift towards Open Access for academic publishing, which raises important questions about how the legitimate rights of publishers are interpreted in legal deposit. Second, regulatory

exceptions that support data-driven research are incompatible with an access model that privileges reading above other forms of access. Third, the replicability of digital materials means that NPLD stops only the legal deposit libraries in making certain materials available; non-governmental organisations have already provided innovative services that make huge swathes of contemporary materials available.

Given these overlapping points, the current restrictive access arrangements only limit access for users of legal deposit libraries, thereby eroding the ability of these libraries to provide innovative services. This creates a gap between trusted repositories, which are tied to legal deposit regulations, and non-governmental organisations. Our proposed 'datafication of the legal deposit library' is a response to these problems that recognises the changing nature of our cultural outputs, while understanding the fundamental importance of reference-only access to legal deposit collections. This proposal would be supported by two complementary priorities for legal deposit: first, closer alignment with copyright regulations, to support reading rooms to develop innovative services; and second, allowing differentiation between OA and non-OA materials to replicate the use originally intended by rights holders. The importance of these suggestions lies in the fact that NPLD collections promise to contain unique materials in quantities that are unprecedented in the history of legal deposit, and that libraries will thus become the sole source for access.

The deluge noted by Francis Bacon was the rapid growth of print sources following the invention of the printing press. To the world after the digital turn, however, it evokes images of information overload through online proliferation. This proliferation has changed user expectations, scholarly publishing models, and primary regulations; it has also allowed data-driven research and re-centralised the importance of understanding the textuality of our cultural sources. We must respond by interrogating legal deposit in a new light, so that our non-print collections are not permanently affected by the legacies of print media.

References

Bodleian Libraries (2017) *History of the Bodleian*. Available at: https://www.bodleian.ox.ac.uk/bodley/about-us/history (Accessed: 3 May 2017).

Brazier, C. (2016) 'Great Libraries? Good Libraries? Digital Collection Development and What it Means for Our Great Research Collections', in Baker, D. and Evans, W. (eds) *Digital Information Strategies: From Applications and Content to Libraries and People*. Waltham, MA: Chandos Publishing, pp. 41–56.

British Library (1998) *Report of the Working Party on Legal Deposit*. Available at: http://www.bl.uk/aboutus/stratpolprog/legaldep/report/index.html (Accessed: 22 October 2018).

British Library (2013) *Security for Electronic Publications*. Available at: https://www.bl.uk/aboutus/legaldeposit/websites/security/ (Accessed: 24 September 2018).

Brooks, M., Heller, M. and Phetteplace, E. (2013) 'Library Labs', *Reference & User Services Quarterly*, 52(3), pp. 186–190.

Brügger, N. (2012) 'Web History and the Web as a Historical Source', *Studies in Contemporary History*, 2. Available at: http://www.zeithistorische-forschungen.de/site/40209295/default.aspx (Accessed: 9 January 2017).

Coyle, K. (2017) 'Creating the Catalog, Before and After FRBR'. *Encuentro di Catalogacion y Metadatos*, Universidad Nacional Autonoma de Mexico, 9 December. Available at: http://kcoyle.net/mexico.html (Accessed: 17 October 2017).

De Beer, M. *et al.* (2016) 'Legal Deposit of Electronic Books - A Review of Challenges Faced by National Libraries', *Library Hi Tech*, 34(1), pp. 87–103.

Department for Digital, Culture, Media & Sport (2019) *Post-Implementation Review of the Legal Deposit Libraries (Non-Print Works) Regulations 2013*. Department for Digital, Culture, Media & Sport. Available at: https://www.gov.uk/government/publications/post-implementation-review-of-the-legal-deposit-libraries-non-print-works-regulations-2013 (Accessed: 11 April 2019).

Eve, M. (2016) *The UK copyright exemption for text and data mining vs. the DMCA and EUCD, Martin Paul Eve*. Available at: https://www.martineve.com/2016/01/07/the-uk-copyright-exemption-for-text-and-data-mining-vs-the-dmca/ (Accessed: 24 October 2018).

Eve, M. P. (2014) *Open Access and the Humanities: Contexts, Controversies and the Future*. Cambridge: Cambridge University Press.

Gantz, J. and Reinsel, D. (2012) *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. Framingham, MA: IDC Go-to-Market Services.

Gibby, R. and Brazier, C. (2012) 'Observations on the Development of Non-Print Legal Deposit in the UK', *Library Review*, 61(5), pp. 362–377.

Gooding, P. (2016) ‘Exploring the information behaviour of users of Welsh Newspapers Online through web log analysis’, *Journal of Documentation*, 72(2), pp. 232–246.

Gooding, P. (2017) *Historic Newspapers in the Digital Age: ‘Search All About It’*. Oxon: Routledge.

Gooding, P., Terras, M. and Berube, L. (2018) ‘Legal Deposit Web Archives and the Digital Humanities: A Universe of Lost Opportunity?’, in. *Digital Humanities 2018*, Mexico City. doi: http://doi.org/10.5281/zenodo.1300011.

Gooding, P., Terras, M. and Berube, L. (2019) *Towards User-Centric Evaluation of Non-Print Legal Deposit: A Digital Library Futures White Paper*. Glasgow, Edinburgh and Norwich: University of Glasgow. Available at: elegaldeposit.org/resources (Accessed: 30 May 2019).

Hale, S. A., Blank, G. and Alexander, V. D. (2017) ‘Live Versus Archive: Comparing a Web Archive to a Population of Web Pages’, in Brügger, N. and Schroeder, R. (eds) *The Web as History: Using Web Archives to Understand the Past and Present*. London: UCL Press. Available at: http://discovery.ucl.ac.uk/1542998/1/The-Web-as-History.pdf (Accessed: 26 September 2018).

Hayles, N. K. (2005) *My mother was a computer: digital subjects and literary texts*. Chicago: University of Chicago Press.

HC Deb (1997) *Non-Print Publications (Legal Deposit)*. Vol. 290, Col. 15563. Available at: https://hansard.parliament.uk/Commons/1997-02-11/debates/609493dd-462b-4560-9110-3cf9f53c7aec/Non-PrintPublications(LegalDeposit)?highlight=%22legal%20deposit%22#contribution-29c1ed93-bf71-4c27-81e7-066affed5975.

HL Deb (2013). Available at: https://hansard.parliament.uk/Lords/2013-03-06/debates/13030683000250/LegalDepositLibraries(Non-PrintWorks)Regulations2013?highlight=%22legal%20deposit%22#contribution-13030683000082.

Hockx-Yu, H. (2016) *Web Archiving at National Libraries: Findings of Stakeholders’ Consultation by the Internet Archive*. Internet Archive. Available at: https://ia800405.us.archive.org/12/items/InternetArchiveStakeholdersConsultationFindingsPublic/InternetArchiveStakeholdersConsultation-Findings_Public.pdf (Accessed: 26 September 2018).

Intellectual Property Office (2014) *Exceptions to Copyright: Research, UK Government*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf.

Kahle, B. (2007) ‘Universal Access to All Knowledge’, *The American Archivist*, 70(1), pp. 23–31.

Kahle, B. (2017) ‘Books from 1923 to 1941 Now Liberated!’, *Internet Archive Blogs*. Available at: http://blog.archive.org/2017/10/10/books-from-1923-to-1941-now-liberated/ (Accessed: 17 October 2017).

Kremerskothen, K. (2013) 'Welcome the British Library to the Commons!', *Flickr Blog*, 16 December. Available at: http://blog.flickr.net/2013/12/16/welcome-the-british-library-to-the-commons/ (Accessed: 26 September 2018).

Lariviere, J. (2000) *Guidelines for Legal Deposit Legislation*. Paris: United Nations Educational, Scientific and Cultural Organization. Available at: http://unesdoc.unesco.org/images/0012/001214/121413eo.pdf (Accessed: 28 June 2017).

Larsen, S. (2005) 'Preserving the Digital Heritage: New Legal Deposit Act in Denmark', *Alexandria: The Journal of National and International Library and Information Issues*. Available at: http://journals.sagepub.com/doi/abs/10.1177/095574900501700204 (Accessed: 5 April 2017).

Leetaru, K. (2017) 'Common Crawl and Unlocking Web Archives for Research', *Forbes*, 28 September. Available at: https://www.forbes.com/sites/kalevleetaru/2017/09/28/common-crawl-and-unlocking-web-archives-for-research/#86131fc3b833 (Accessed: 7 November 2017).

Legal Deposit Libraries Act 2003 c.28 (2003). Available at: https://www.legislation.gov.uk/ukpga/2003/28/contents (Accessed: 16 April 2019).

Library of Congress (2009) *WARC, Web ARChive file format*. Available at: https://www.loc.gov/preservation/digital/formats/fdd/fdd000236.shtml (Accessed: 24 October 2018).

Mahey, M. (2018) *Building Library Labs around the world - the event and complete our survey! - Digital scholarship blog*. Available at: https://blogs.bl.uk/digital-scholarship/2018/09/building-library-labs-around-the-world.html (Accessed: 25 October 2018).

Manovich, L. (2001) *The Language of New Media*. Boston, MA: MIT Press.

McGregor, N. and Farquhar, A. (2013) 'The Digital Scholarship Training Programme at British Library', in. *Digital Humanities 2013*, Lincoln. Available at: http://dh2013.unl.edu/abstracts/ab-264.html (Accessed: 21 November 2013).

Menell, P. S. (2007) 'Knowledge Accessibility and Preservation Policy for the Digital Age', *Houston Law Review*, 44(4). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=999801 (Accessed: 11 July 2017).

Meyer, E. T. *et al.* (2017) 'Analysing the UK Web Domain and Exploring 15 Years of UK Universities on the Web', in Brügger, N. and Schroeder, R. (eds) *The Web as History: Using Web Archives to Understand the Past and Present*. London: UCL Press. Available at: http://discovery.ucl.ac.uk/1542998/1/The-Web-as-History.pdf (Accessed: 26 September 2018).

Milligan, I. (2016) 'Lost in the Infinite Archive: The Promise and Pitfalls of Web Archives', *International Journal of Humanities and Arts Computing*, 10(1), pp. 78–94. doi: 10.3366/ijhac.2016.0161.

National Library of Scotland (2013) *Electronic Legal Deposit*. Available at: http://www.nls.uk/news/press/2013/04/electronic-legal-deposit (Accessed: 14 August 2013).

National Library of Scotland (2016) *National Library Opens in Glasgow*. Available at: https://www.nls.uk/news/press/2016/09/library-opens-at-kelvin-hall (Accessed: 12 September 2017).

National Library of Scotland (2019) *Launch of Data Foundry website - National Library of Scotland*. Available at: https://www.nls.uk/news/archive/2019/09/data-foundry (Accessed: 26 November 2019).

Padilla, T. *et al.* (2019) 'Final Report - Always Already Computational: Collections as Data'. doi: 10.5281/zenodo.3152935.

Philip, I. (1983) *The Bodleian Library in the Seventeenth and Eighteenth Centuries*. Oxford: Clarendon Press.

Pugh, S. L. (2010) 'The National Library of Wales and Flickr Commons', *Cultural Heritage: A UKOLN Blog for the Cultural Heritage Sector*, 26 July. Available at: http://blogs.ukoln.ac.uk/cultural-heritage/2010/07/26/the-national-library-of-wales-and-flickr-commons/ (Accessed: 11 January 2014).

Quint, B. (2001) 'Don't Burn Books! Burn Librarians!! A Review of Nicholson Baker's Double Fold: Libraries and the Assault on Paper', *Searcher*, 9(6). Available at: http://www.infotoday.com/searcher/jun01/voice.htm.

Rossi, A. (2012) '80 Terabytes of Archived Web Crawl Data Available for Research', *Internet Archive Blogs*, 26 October. Available at: http://blog.archive.org/2012/10/26/80-terabytes-of-archived-web-crawl-data-available-for-research/ (Accessed: 26 September 2018).

Scholz, H. and Miles, T. (2015) 'From the British Library to Flickr to Europeana - 60000 public domain images on the move', *Europeana Pro*. Available at: https://pro.europeana.eu/post/from-the-british-library-to-flickr-to-europeana-60000-public-dom (Accessed: 25 October 2018).

Science Europe (2018) 'What is cOALition S?', *Science Europe: Shaping the Future of Research*, 4 September. Available at: https://www.scienceeurope.org/coalition-s/ (Accessed: 19 September 2018).

Seadle, M. (2001) 'Copyright in the Networked World: Digital Legal Deposit', *Library Hi Tech*, 18(3), pp. 299–303.

Spedding, J. (1861) *The Letters and the Life of Francis Bacon*. London.

Tanner, S. (2013) 'The Value of Welsh Newspapers Online', *When the Data Hits the Fan: The Blog of Simon Tanner*, 28 March. Available at: http://simon-tanner.blogspot.co.uk/2013/03/the-value-of-welsh-newspapers-online.html (Accessed: 28 March 2013).

The Legal Deposit Libraries (Non-Print Works) Regulations 2013 (2013). Available at: http://www.legislation.gov.uk/uksi/2013/777/contents/made (Accessed: 15 August 2013).

Townsend Gard, E. (2017) 'Creating a Last Twenty (L20) Collection: Implementing Section 108(h) in Libraries, Archives and Museums', *Libraries, Archives and Museums*. doi: 10.2139/ssrn.3049158.

UK Research and Innovation (2013) *RCUK Policy on Open Access and Supporting Guidance*. Available at: https://www.ukri.org/files/legacy/documents/rcukopenaccesspolicy-pdf/ (Accessed: 24 September 2018).

Warwick, C. *et al.* (2008) 'If You Build It Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities', *Literary and Linguistic Computing*, 23(1), pp. 85–102.

Weber, M. S. (2017) 'The Tumultuous History of News on the Web', in Brügger, N. and Schroeder, R. (eds) *The Web as History: Using Web Archives to Understand the Past and Present*. London: UCL Press. Available at: http://discovery.ucl.ac.uk/1542998/1/The-Web-as-History.pdf (Accessed: 26 September 2018).

Wilkinson, J. M., Pollard, T. and Farquhar, A. (2010) 'British Library Dataset Programme: Supporting Research in the Library of the 21st Century', *LIBER Quarterly*, 20(1), pp. 94–104. doi: 10.18352/lq.7979.

Winters, J. (2017) 'Coda: Web Archives for Humanities Research - Some Reflections', in *The Web as History*. London: UCL Press, pp. 238–248.

[i] See the chapter in this volume by Linda Arnold-Stratford and Richard Ovenden for further elaboration on the development and implementation of access arrangements for Non-Print Legal Deposit in the United Kingdom.

[ii] For further details of the technical implementation, see Arnold-Stratford and Ovenden in this volume.

[iii] The Digital Library Futures White Paper (Gooding, Terras and Berube, 2019, p. 13) contains a photo from the Cambridge University Library, which demonstrates the appearance and layout of the NPLD terminals in the Library.

[iv] The Bill that Kahle refers to is the Copyright Term Extension Act (CTEA) of 1998, which extended copyright to the life of the author plus 70 years. It was informally named after the deceased singer and congressman Sonny Bono, who was one of twelve sponsors of a similar bill. High profile lobbyists also included Disney, leading it to be derisively billed 'The Mickey Mouse Protection Act,' due to the proximity of the copyright term extension to the point at which the rights to Mickey Mouse would have entered the public domain.

[v] WARC is the Web ARCHive File Format, which specifies a method for combining multiple digital resources into an aggregate archival file with related information. It consists of metadata fields to support the retrieval of each harvested resource, and allows the storage of content blocks to store resources in any format: this can include binary image or audio-visual files that may be embedded in HTML pages (Library of Congress, 2009).

[vi] While at first glance this looks like a spelling mistake, cOAlition S is the official name of a declaration of commitment to Open Access. Further details are available at https://www.scienceeurope.org/coalition-s/.

[vii] The issue of perpetual copyright was, for instance, brought up as a key concern by the library submission to the recent Post-Implementation Review of Non-Print Legal Deposit. For further discussion see Annex A of the DCMS review (Department for Digital, Culture, Media & Sport, 2019)