

---

# DIGITAL PRESERVATION STRATEGY AND PRESERVATION PLANNING AT THE BRITISH LIBRARY

---

## Michael DAY

Digital Preservation Manager, The British Library

- The British Library is the custodian of diverse and rapidly-growing collections of digital content, both born-digital and digitised. Its recently published vision document, *Living Knowledge*, has confirmed that custodianship is the Library's "first and core purpose", the one on which all of its other purposes depend. The strategy also specifically highlighted the need for the Library to address the challenges of the long-term preservation of digital content, in particular for the born-digital content received through legal deposit, which includes the periodic (at least annual) capture of the entire UK Web domain. This paper explores some of the activities being undertaken by the Library's digital preservation team as it seeks to implement a preservation planning capacity, focusing on the initial capture of preservation requirements through collection profiling and a programme of file format assessments.
- La British Library est dépositaire de collections de contenus numériques, diverses et en rapide expansion, tant numérisées que sous forme numérique d'origine. Elle a récemment publié un document à vision stratégique, intitulé "Living Knowledge" (connaissance vivante), confirmant que cette mission de conservation est la mission première et fondamentale de la Bibliothèque, celle sur laquelle reposent toutes les autres missions. Cette stratégie a également spécifiquement mis en évidence la nécessité pour la Bibliothèque de prendre en considération les défis posés par la conservation à long terme des contenus numériques, en particulier les contenus numériques d'origine reçus au travers du dépôt légal, qui inclut la capture périodique (au minimum annuelle) de l'ensemble du domaine web du Royaume-Uni. Le présent article explore quelques-unes des activités entreprises par l'équipe de la Bibliothèque chargée de la préservation digitale, dans sa recherche pour implémenter une capacité de planification en vue de cette préservation, en mettant l'accent sur la saisie initiale des exigences de préservation au travers du profilage des collections et sur un programme d'évaluation des formats de fichiers.
- British Library is de hoeder van diverse en snel groeiende verzamelingen van digitale content, van zowel born-digital als gedigitaliseerde documenten. Zijn onlangs verschenen visiedocument "Living Knowledge" bevestigt dat de bewaarfunctie van de bibliotheek primordiaal is en de core business blijft waarvan de andere doelstellingen afgeleid worden. De strategie beklemtoont de noodzaak van British Library om een bewaring op lange termijn van digitale stukken uit te werken, vooral dan wat betreft de born-digital content afkomstig uit het wettelijk depot, dit inclusief het periodiek (ten minste jaarlijks) vastleggen van het gehele Britse webdomein. Het artikel onderzoekt een deel van de activiteiten uitgevoerd door het digital preservation team in zijn streven om voor deze bewaring een capaciteitsplanning te realiseren via het focussen op de initiële bewaringseisen dit via een collectieprofilering en een programma van bestandsformaatevaluaties.

The British Library is the custodian of ever-increasing amounts of digital content, both born-digital and digitised. As a national library, it is the Library's responsibility to preserve this content and make it available for current and future users. The preservation of digital content, is not straightforward, however, in that it typically requires actions and interventions throughout content lifecycles, often far earlier and far more frequently than would be typical with physical collections<sup>1</sup>. Above all, choices need to be made about which preservation strategies to adopt, which are currently mostly variations upon emulation or content migration<sup>2</sup>. Regardless of preservation strategy, however, decisions need to be made about target formats or operating environments, as well as on the tools and software required to

manage digital collections over the long-term, e.g. for format identification and validation, migration, integrity checking, preservation assessment, etc. The process of making and recording these decisions is broadly known as preservation planning<sup>3</sup>.

Preservation planning depends upon the production of preservation plans, understood by Becker, et al. as action plans for "preserving a specific set of objects for a given purpose".<sup>4</sup> A typical preservation plan would need to take into account, for example, organisational contexts and technical environments, as well as the preservation objectives and requirements of particular collections. It would also need to justify any decisions taken and specify in more detail the series of activities needed to preserve the collection (the latter is often known

as a preservation action plan)<sup>5</sup>. The importance of preservation planning for digital preservation is indicated by its inclusion in emerging standards for repository audit and certification, e.g. ISO 16363:2012<sup>6</sup>.

The British Library is currently in the process of developing its preservation planning capacity. This is being done both pragmatically and incrementally. After providing an overview of the Library and its digital collections, this paper will introduce some of the initial steps undertaken by the Library's digital preservation team in implementing a preservation planning capacity, focusing on the initial capture of preservation requirements through collection profiling and a programme of file format assessments.

## The British Library

The British Library is the UK's national library. It was established on 1 July 1973 and defined in legislation as "a national centre for reference, study and bibliographical and other information services, in relation both to scientific and technological matters and to the humanities."<sup>7</sup> The Library inherited much of its national library role – including its status as a legal deposit library – from the British Museum, whose printed book and manuscript collections provided the core of the Library's initial holdings. These have been supplemented since by the incorporation of other collections, including the India Office Library and Records and the National Sound Archive. The British Library is now one of the largest research libraries in the world, with significant collections of printed works, manuscripts, maps, sound recordings, postage stamps, and much else. The Library is the custodian of a great many unique items, including two of the four surviving copies of the 1215 Magna Carta manuscript, which was the focus of a major exhibition held at the Library in 2015<sup>8</sup>. The British Library is based in two main locations. Firstly, there is a purpose-built library building at St Pancras in central London – now part of a fast-growing "Knowledge Quarter" in the capital. This building was designed by Colin St John Wilson and built between 1982 and 1997; it was awarded grade I listed building status by Historic England in August 2015. Secondly, there is a site at Boston Spa in Yorkshire, which provides the hub for the Library's document supply serviceWs, as well as having significant physical storage capacity, including a dedicated Newspaper Storage Building that was opened in 2013<sup>9</sup>. Living Knowledge

In January 2015, the British Library published a new vision document to cover the development

of the Library up to its 50th anniversary in 2023. Living Knowledge: The British Library, 2015-2023 outlined six main purposes of the British Library<sup>10</sup>:

- Custodianship – We build, curate and preserve the UK's national collection of published, written and digital content;
- Research – We support and stimulate research of all kinds;
- Business – We help businesses to innovate and grow;
- Culture – We engage everyone with memorable cultural experiences;
- Learning – We inspire young people and learners of all ages;
- International – We work with partners around the world to advance knowledge and mutual understanding.

Living Knowledge reaffirmed that *custodianship* was the British Library's "*first and core purpose, the one on which all the others depend.*" In undertaking custodianship, it noted the Library's dependence on a wide range of specialist skills, covering collection management (e.g. content processing, metadata, conservation and preservation) as well as scholarly and curatorial expertise. The document also highlighted a few specific custodianship priorities that would need to be dealt with over the next few years. One of these was to address the preservation of audio and recorded music collections, where there was a concern about the longevity of the physical formats on which sound recordings were kept. This is already being addressed by the Save our Sounds project, initiated by a grant of £9.5 million from the UK Heritage Lottery Fund, announced in May 2015. Another custodianship priority included in Living Knowledge outlined the need to address the long-term preservation of born-digital content received through legal deposit. It described the recent extension of legal deposit to non-print content, including the periodic capture of the entire UK Web domain, as "the heart of our single greatest endeavour in digital custodianship".

Begun in April 2013, we are at the earliest stages of a journey that is set, over time, to transform what it means to be a national memory institution. Our challenges for the next phase of the project are to develop the scale and accessibility of the collection, and to ensure a proper level of investment in its future storage and preservation. (p. 13)

Living Knowledge acknowledges the extent to which the British Library is already a digital library.

## Digital collections

The British Library's collections contain both born-digital and digitised content.

The growth in born-digital content to some extent simply reflects changes in the publishing industry, but, as Living Knowledge has noted, this has been amplified since April 2013 by the extension of UK legal deposit provisions to non-print content, including the Web. As a legal deposit library, the British Library inherited from the British Museum the right to receive copies of certain types of printed content published in the UK, including books, newspapers, printed music and maps (a right that ultimately dates back to the seventeenth century). Primary legislation supporting the extension of legal deposit to non-print content in the UK was passed in 2003. After a decade of planning and negotiation, the Legal Deposit Libraries (Non-Print Works) Regulations 2013 came into force on the 6th April 2013<sup>13</sup>. This, for the first time, enabled the British Library and the other UK copyright libraries to claim certain classes of non-print content under legal deposit provisions and to make it available to on-site users. In the British Library, this has led to the development of specialised ingest workflows to receive and process electronic journals and books, as well as a massive scaling-up of the Library's Web archiving activities to include the periodic (at least annual) capture of the entire UK Web Domain. The first of these domain crawls (running from April to June 2013) captured approximately 31 TB of compressed data<sup>15</sup>. Other born-digital content received by the Library includes geospatial datasets received from the Ordnance Survey and the Ordnance Survey of Northern Ireland. These include, from 2007, annual snapshots of Ordnance Survey MasterMap™ data, which is based on the Geography Markup Language (GML).

At the same time, the British Library has also spent a great deal of time exploring how to make elements of its existing collections available to wider audiences through digitisation and other digital scholarship initiatives. The Library's involvement in digitisation programmes started in the 1990s with some relatively small-scale projects undertaken as part of its Initiatives for Access programme. For example, the Electronic Beowulf project produced a facsimile edition of the Early English poem based on high-resolution images of the Library's version of the manuscript, linked to textual transcriptions made over several centuries and other critical apparatus<sup>18</sup>. This was initially published on CD-ROM in 2000, although the most recent version, Electronic Beowulf 4.0, is now a freely-available online resource. Since the 1990s,

the Library has scaled-up its digitisation activities considerably. The Library continues to digitise selected manuscripts and it makes many of them freely available through its Digitised Manuscripts website or (selectively) through online resources like Discovering Literature. The Library also undertakes much digitisation activity as part of funded grant programmes or in partnership with the commercial sector. For example, the Library has used grant funding to digitise many sound recordings and, wherever possible, it makes the results available through the British Library Sounds webpage; an activity that will be scaled up as part of the Save our Sounds project. Also, in the past five years, as part of a broader European collaboration linked to the centenary of the outbreak of the First World War, a range of different content types (monographs, photographs, archives, musical scores, etc.) were digitised and made available through the Europeana 1914-1918 portal. The Library's partnerships with the commercial sector (mainly secondary publishers) has also allowed it to experiment with large-scale digitisation, current examples being a partnership with Findmypast on the British Newspaper Archive and the Library's participation in the Google Books program.

The end result is that the British Library is the custodian of diverse and rapidly growing collections of digital content, both born-digital and digitised.

## Digital preservation at the British Library

Given its responsibility to preserve this content and to make it available to current and future users, the British Library has, over the past few decades, made significant progress in responding to the organisational and infrastructural challenges of digital preservation.

### Digital repository

A key part of the British Library's infrastructure is its digital repository, known as the Digital Library System (DLS). Prior to the extension of legal deposit to non-print content in 2013, the Library and the other UK Legal Deposit Libraries invested heavily in developing scalable solutions to the acquisition, storage and management of very large amounts of digital content. The repository has been described as a "single location to ingest, store, preserve, manage, discover and provide controlled access to digital content assets."<sup>19</sup> While designed as an integrated storage system, it has been implemented in a highly distributed way, with content replicated in four geographically separated storage nodes

(based in London, Boston Spa, Edinburgh and Aberystwyth), with additional access gateways at two of the other legal deposit libraries (Oxford and Cambridge).

Some practical features of the British Library's repository have been described in an APARSEN project deliverable<sup>20</sup>. Ingest usually takes place at one of the British Library's sites, with different ingest streams defined for different types of content, e.g. e-journals, digitized newspapers, or Web archive content. All objects have a signature file, which includes a hash value and timestamp, and content is automatically replicated on all four of the storage nodes after ingest. The system assumes that some bit-loss is inevitable when storing extremely large amounts of data. The repository has, therefore, been designed to be self-checking and self-healing; there are periodic integrity checks, and if an object is found to be damaged, it is replaced by a good copy from another node. The repository is designed to be scalable and vendor-independent, using commodity hardware that can be added to as required.

## Digital preservation strategy

The British Library's Digital Preservation Strategy, 2013-2016 starts with the assumption that it is the Library's responsibility to preserve and make available content to current and future users. The strategy itself, which was approved in March 2013, outlined four main strategic priorities, i.e. to<sup>21</sup>:

- Ensure our digital repository can store and preserve our collections for the long term;
- Manage the risks and challenges associated with digital preservation throughout the digital collection content lifecycle;
- Embed digital sustainability as an organisational principle for digital library planning and development;
- Benefit from collaboration with other national and international institutions on digital preservation initiatives.

Fundamental to the implementation of this strategy is a sound understanding of the Library's digital collections and the preservation challenges that underpin them. The gathering of preservation requirements provides a baseline for the Library's development of a preservation planning capacity.

## Preservation planning

Preservation planning is an essential part of any successful digital preservation strategy because it reflects the need for decisions to be made about appropriate preservation approaches, tools and formats at multiple points within content lifecycles.

Preservation planning potentially covers a very wide range of activities, but the starting point – as with so much else in digital preservation – has usually been the Reference Model for an Open Archival Information System (OAIS), which has also been approved as an international standard (ISO 14721:2012)<sup>22</sup>. The OAIS Model defines a functional entity for Preservation Planning, that would provide "the services and functions for monitoring the environment of the OAIS [i.e. an archive, broadly defined], providing recommendations and preservation plans to ensure that the information stored in the OAIS remains accessible to, and understandable by, the Designated Community [i.e. the expected users] over the Long Term, even if the original computing environment becomes obsolete."<sup>23</sup> The model also provides some specific examples of what functions might be required<sup>24</sup>:

Preservation Planning functions include evaluating the contents of the Archive and periodically recommending archival information updates, recommending the migration of current Archive holdings, developing recommendations for Archive standards and policies, providing periodic risk analysis reports, and monitoring changes in the technology environment and in the Designated Community's service requirements and Knowledge Base. [...] Preservation Planning also develops detailed Migration plans, software prototypes and test plans to enable implementation of Administration migration goals. (p. 1-14)

Following the OAIS Reference Model, the concept of preservation planning has also started to be embedded in the emerging standards being developed to support repository audit and certification. For example, Section 4.3 in the ISO 16363:2012 Audit and Certification of Trustworthy Digital Repositories standard includes four separate requirements relating to preservation planning<sup>25</sup>:

- 4.3.1 The repository shall have documented preservation strategies relevant to its holdings;
- 4.3.2 The repository shall have mechanisms in place for monitoring its preservation environment;

- 4.3.3 The repository shall have mechanisms to change its preservation plans as a result of its monitoring activities;
- 4.3.4 The repository shall provide evidence of the effectiveness of its preservation activities. (p. 4-16 ff.)

Perhaps the most comprehensive attempt so far to develop a structured approach to preservation planning has been the Plato decision-support tool, initially developed as part of the European Union-funded Planets (Preservation and Long-term Access through Networked Services) project<sup>26</sup>. Plato provides a methodology and a software tool to support the systematic capture of preservation requirements from various stakeholders and then to match these to potential preservation strategies for further analysis. The result of this is a recommendation that can form the basis of a preservation plan, which contains information on contexts as well as recording the evidence base underpinning the decision. Several attempts have been made to integrate Plato with other digital preservation tools. For example, researchers from the KeepIt and Planets projects integrated Plato and other digital preservation tools with the ePrints repository software, creating plugins to ePrints that would support the development of preservation workflows, including the generation of preservation plans and action plans<sup>27</sup>. Also, the European Union-funded SCAPE (Scalable Preservation Environments) project also explored how best to integrate Plato with a range of other digital preservation tools and services<sup>28</sup>. In terms SCAPE, it was intended that the resulting preservation plans would be able to document collections, their institutional context, and all of the decision-making process that led to the selection of a particular preservation action. It would also contain a preservation action plan that would contain all of the information necessary to apply the preservation action as well as an executable action plan that could be deployed through a workflow management system like Taverna<sup>29</sup>.

## Preservation planning at the British Library

The British Library is still very much in the process of developing its preservation planning capacity. However, its digital preservation team has begun the process of developing a deeper understanding of the Library's collections and their preservation requirements. Two outcomes of this approach have been the development of collection profiles and a programme of file format assessments.

## Collection profiling

Collection profiling was an attempt to work with curators and other content specialists to develop descriptive profiles of the British Library's high-level digital collection areas, with the aim of capturing key knowledge about the collections and their specific preservation requirements<sup>30</sup>.

The British Library's collection profiling activity drew its inspiration from other content profiling activities, largely those based on a structured dialogue with curators and other content specialists. Particularly influential was the work of the National Library of Australia (NLA) on identifying 'preservation intent' for their digital collections<sup>31</sup>. As part of their approach to preservation planning, digital preservation specialists at the NLA consulted with content specialists (collection managers, curators) in order to develop 'plain-language' statements about "which collection materials, and which copies of collection materials, need to remain accessible for an extended period, and which ones can be discarded when no longer in use or when access to them becomes troublesome." Content specialists at the NLA were also "asked to make broad statements clarifying what 'accessible' means by stating the priority elements that need to be re-presented in any future access for each kind of digital object type in their collections." This both became a means of ensuring that curators and other collection specialists take responsibility for deciding what will happen to collections and is essential for preservation planning. Webb, et al. write that "without it, we are left floundering between assumptions that every characteristic of every digital item has to be maintained forever (almost certainly an impossible expectation) and assumptions that it is good enough to store data safely and let future users worry about how to access it (almost certainly an inadequate response)." Capturing elements of what the NLA refers to as preservation intent seemed vital for the success of the British Library's collection profiling activity.

The digital preservation team also looked at a number of other content-based profile initiatives, in particular the Digital Content Reviews (DCR) for Life Cycle Management developed by MIT Libraries and the Data Curation Profiles developed by Purdue University Libraries. Purdue's Data Curation Profiles are a means of capturing basic information about research datasets in order to support their curation and reuse. The profile provided a comprehensive framework (an interview structure) that could be used to gather information about datasets and their potential

re-use. Once completed, profiles would then be able to help guide decision-making about the management of datasets, as well as inform those providing research data management services of any specific requirements<sup>32</sup>. While the Purdue team's framework was probably too focused on one particular type of content to be useful for the British Library's immediate purposes, the general approach clearly demonstrated the benefits of using content profiles to support lifecycle management. The MIT Libraries Digital Content Reviews for Life Cycle Management took a similar lifecycle management view to the Purdue team, but – more in common with the emerging British Library profiles – they were primarily intended to help capture information about the implications of collecting certain types of digital content<sup>33</sup>.

The development of collection profiles at the British Library was broken down into a number of smaller stages. The initial tasks were to identify the British Library's high-level digital collection areas and to develop an initial template to capture the required information<sup>34</sup>. Identifying high-level collections was not easy, as there was no agreed list of digital collection types held by the Library. Those lists that did exist - e.g. those provided by the catalogue or website - often included, for reasons of practicality, content types at several different levels of granularity. The initial list of high-level collection types included some that were firmly based on resource type (e.g. sound, multimedia), others that were multi-faceted but based on particular content streams (e.g. Web archives); and others that followed more traditional categorizations of library collections, updated for the digital era (e.g. journals, books). Experience with developing the profiles led us to take a slightly more pragmatic approach, largely because we were not trying to produce a definitive taxonomy of all digital collection types held by the Library, but simply to be able to identify collections at a sufficient (and logical) level of granularity in order to get started on the development of content profiles.

Developing the profiles themselves was also an excellent opportunity to develop conversations with curators and other content specialists on identifying the specific preservation requirements of the British Library's digital collections. As a contribution to the development of a more comprehensive preservation planning function, this has had a number of benefits:

- The massive scale of content held by the British Library means that profiling can be undertaken at collection level, hopefully simplifying the identification of preservation

requirements, and the tools necessary to facilitate these;

- Collection profiling opens a forum in which all collection stakeholders (the people who make decisions at different lifecycle stages) can express the challenges faced by specific content types. This will help with the development of a shared understanding of digital preservation requirements from both curatorial and technical perspectives;
- Corporate understanding of the collections held by the British Library is enriched through the sharing of information about collections between the different departments which make collection management decisions.

The British Library now has a set of collection profiles that can inform the further development of its preservation planning capability and provide a baseline for future requirements gathering.

### File format assessments

Another strand of work led by the digital preservation team has been based on the assessment of file formats<sup>35</sup>. File formats are problematic in digital preservation for a variety of reasons, but it was felt that some of the criteria used to assess file formats were not always consistent with practical experience. It has been noted that working with formats often reveal significant challenges that are not always apparent in assessments made from mainly theoretical perspectives<sup>36</sup>.

When developing a methodology for completing file format assessments for selected formats used by the British Library, the digital preservation team was keen to base its assessments on evidential principles. In particular, the team was determined to move away from numerical scoring based approaches or simple measures based on the availability or completeness of documentation. In addition, any recommendation to choose a preservation format different from that in which the content was received needed to be supported by strong empirical evidence of the risks and benefits involved.

The file format assessments produced by the British Library aim simply to provide evidence-based recommendations around the use of that format. Any potential risks in adopting the format are identified, noting any possible mitigation. Format assessments have been produced so far for a number of formats of interest to the Library, including: TIFF (Tagged Image File Format), JPEG 2000 (Joint Photographic Experts Group), PDF (Portable Document Format), XML (Extensible

Markup Language), EPUB, JATS (Journal Archiving and Interchange Tag Suite), NTF (National Transfer Format), and WAVE (Waveform Audio File Format). Public versions of the assessments are available from the Digital Preservation Coalition wiki<sup>37</sup>. The results from these assessments will, in turn, feed into the evidence base that is required to support successful preservation planning.

As a national library, the British Library has a responsibility to preserve its ever growing digital collections and to make them available to current and future users. The extension of UK legal deposit provisions to non-print works in April 2013 represented a major milestone in the Library's acceptance of its role as a digital library. Over the past decade or so, the Library has made significant progress in responding to the organisational and infrastructural challenges of digital preservation, not least in the establishment of infrastructures

to support the processing and storage of digital content as well as in the development of a digital preservation strategy. The next steps for the British Library will be to develop a comprehensive preservation planning capacity. The Library's digital preservation team have made a start on this by developing a deeper understanding of the collections and their curatorial and technical context.

**Michael Day**

*The British Library*  
96 Euston Rd  
London NW1 2DB  
United Kingdom  
<http://www.bl.uk>  
[michael.day@bl.uk](mailto:michael.day@bl.uk)

February 2016

## Notes

1. British Library. *Digital Preservation Strategy, 2013-2016* [on line] <<http://www.bl.uk/aboutus/stratpolprog/collectioncare/digitalpreservation/strategy/dpstrategy.html>>. The British Library, 2013 (consulted on 17 February 2016).
2. Borghoff, Uwe M.; Rödiger, Peter; Scheffczyk, Jan; Schmitz, Lothar. *Long-term preservation of digital documents*. Springer-Verlag, 2006.
3. CCSDS 650.0-M-2. Reference Model for an Open Archival Information System (OAIS) [on line]. CCSDS Recommended Practice, 2012, p. 1-14. <<http://public.ccsds.org/publications/archive/650x0m2.pdf>> (consulted on 17 February 2016).
4. Becker, Christoph; Kulovits, Hannes; Guttenbrunner, Mark; Strodl, Stephan; Rauber, Andreas; Hofman, Hans; Systematic planning for digital preservation: evaluating potential strategies and building preservation plans. *International Journal on Digital Libraries* [on line], 2009, vol. 10, p. 133-157. <<http://link.springer.com/article/10.1007%2Fs00799-009-0057-1>> (consulted on 17 February 2016).
5. See note 4.
6. ISO 16363:2012. Space data and information transfer systems – Audit and certification of trustworthy digital repositories. International Organization for Standardization, 2012.
7. *British Library Act 1972* [on line]. Her Majesty's Stationery Office, 1972 <<http://www.legislation.gov.uk/ukpga/1972/54/contents>> (consulted on 17 February 2016).
8. Breay, Claire; Harrison, Julian (eds.) *Magna Carta: law, liberty, legacy*. The British Library, 2015.
9. Novotny, Deborah. The changing face of storage at the British Library [on line]. In *Proceedings of the 78th IFLA General Conference and Assembly*. Helsinki, Finland, 11-17 August 2012 <<http://conference.ifla.org/past-wlic/2012/102-novotny-en.pdf>> (consulted on 17 February 2016).
10. British Library. *Living knowledge: The British Library, 2015-2023* [on line]. The British Library, 2015 <<http://www.bl.uk/aboutus/foi/pubsch/pubscheme3/living-knowledge-2015-2023.pdf>> (consulted on 17 February 2016).
11. See note 10.
12. Brazier, Caroline. born.digital@british.library: the opportunities and challenges of implementing a digital collection development strategy [on line]. In *Proceedings of the IFLA World Library and Information Congress*, Singapore, 17-23 August 2013 <<http://library.ifla.org/222/1/198-brazier-en.pdf>> (consulted on 17 February 2016).
13. S.I. 2013 No. 777. *Legal Deposit Libraries (Non-Print Works) Regulations 2013* [on line]. The Stationery Office, 2013 <<http://www.legislation.gov.uk/uksi/2013/777/contents>> (consulted on 17 February 2016).

14. Gibby, Richard; Brazier, Caroline. Observations on the development of non-print legal deposit in the UK. *Library Review* [on line], 2012, vol. 61, no. 5, p. 362-377. <<http://www.emeraldinsight.com/doi/full/10.1108/00242531211280487>> (consulted on 17 February 2016).
15. Webster, Peter. *Crawling the UK web domain* [on line]. UK Web Archive blog, 16 September 2013 <<http://britishlibrary.typepad.co.uk/webarchive/2013/09/domaincrawl.html>> (consulted on 17 February 2016).
16. Fleet, Christopher; Kowal, Kimberly C. Planning the migration to OS MasterMap in the UK Legal Deposit Libraries. *Cartographic Journal* [on line], 2005, vol. 42, no. 3, p. 214-221. <<http://www.tandfonline.com/doi/abs/10.1179/000870405X77147>> (consulted on 17 February 2016).
17. Carpenter, Leona; Shaw, Simon; Prescott, Andrew (eds.) *Towards the digital library: the British Library's Initiatives for Action programme*. The British Library, 1998.
18. Prescott, Andrew. The *Electronic Beowulf* and digital restoration. *Literary and Linguistic Computing*, [on line]. 1997, vol. 12, no. 3, p. 185-196. <[http://www.uky.edu/~kiernan/eBeo\\_archives/articles90s/ajp-llc.htm](http://www.uky.edu/~kiernan/eBeo_archives/articles90s/ajp-llc.htm)> (consulted on 17 February 2016).
19. Fleming, Patrick. The British Library Newspaper Strategy: developing collaboration with publishers to digitise back runs and to ingest born digital newspapers. In: Walravens, Hartmut (ed.) *Newspapers: legal deposit and research in the digital era*. De Gruyter Saur, 2011, p. 21-30.
20. APARSEN. *Storage solutions summary of inputs* [on line]. APARSEN Deliverable D23.1, 2013 <<http://www.alliancepermanentaccess.org/index.php/about-aparsen/aparsen-deliverables/>> (consulted on 17 February 2016).
21. British Library. *Digital Preservation Strategy, 2013-2016* [on line]. The British Library, 2013 <<http://www.bl.uk/aboutus/stratpolprog/collectioncare/digitalpreservation/strategy/dpstrategy.html>> (consulted on 17 February 2016).
22. ISO 14721:2012. Space data and information transfer systems – Open archival information system (OAIS) – Reference model. International Organization for Standardization, 2012.
23. CCSDS 650.0-M-2. Reference Model for an Open Archival Information System (OAIS) [on line]. CCSDS Recommended Practice, 2012 <<http://public.ccsds.org/publications/archive/650x0m2.pdf>> (consulted on 17 February 2016).
24. See note 21.
25. CCSDS 652.0-M-1. Audit and certification of trustworthy digital repositories [on line]. CCSDS Recommended Practice, 2011 <<http://public.ccsds.org/publications/archive/652x0m1.pdf>> (consulted on 17 February 2016).
26. Becker, Christoph; Kulovits, Hannes; Rauber, Andreas; Hofman, Hans. Plato: a service oriented decision support system for preservation planning [on line]. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries*, Pittsburgh, PA, USA, 16-20 June, 2008. ACM Press, 2008, p. 367-370. <<http://dl.acm.org/citation.cfm?doid=1378889.1378954>> (consulted on 17 February 2016).
27. Hitchcock, Steve; Tarrant, David; Carr, Les; Kulovits, Hannes; Rauber, Andreas. Connecting preservation planning and Plato with digital repository interfaces [on line]. In *Proceedings of iPRES 2010, 7th International Conference on Preservation of Digital Objects*, Vienna, Austria, 19-24 September 2010 <<http://www.ifs.tuwien.ac.at/dp/ipres2010/papers/tarrant-65.pdf>> (consulted on 17 February 2016).
28. May, Peter; Wilson, Carl. *Technical architecture report, v2* [on line]. SCAPE Deliverable D2.3, 2014 <<http://www.scape-project.eu/deliverable/d2-3-technical-architecture-report-v2>> (consulted on 17 February 2016).
29. Wolstencroft, Katherine; Haines, Robert; Fellows, Donal; Williams, Alan; Withers, David; Owen, Stuart; Soiland-Reyes, Stian; Dunlop, Ian; Nenadic, Aleksandra; Fisher, Paul; Bhagat, Jiten; Belhajjame, Khalid; Bacall, Finn; Hardisty, Alex; Nieva de la Hidalga, Abraham; Balcazar Vargas, Maria P.; Sufi, Shoaib; Goble, Carole. The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud. *Nucleic Acids Research* [on line]. 2003, vol. 41, no. W1, p. W557-W561. <<http://nar.oxfordjournals.org/content/41/W1/W557>> (consulted on 17 February 2016).
30. Day, Michael; MacDonald, Ann; Kimura, Akiko; Pennock, Maureen. Identifying digital preservation requirements: Digital preservation strategy and collection profiling at the British Library [on line]. In *iPRES 2014: Proceedings of the 11th International Conference on Preservation of Digital Objects 2014*. Melbourne, Australia, 6-10 October 2014 <[https://phaidra.univie.ac.at/detail\\_object/o:378119](https://phaidra.univie.ac.at/detail_object/o:378119)> (consulted on 17 February 2016).
31. Webb, Colin; Pearson, David; Koerbin, Paul. 'Oh, you wanted us to preserve that?!' Statements of preservation intent for the National Library of Australia's digital collections. *D-Lib Magazine* [on line]. 2013, vol. 19, nos. 1/2. <<http://www.dlib.org/dlib/january13/webb/01webb.html>> (consulted on 17 February 2016).
32. Witt, Michael; Carlson, Jacob; Brandt, D. Scott; Cragin, Melissa H.; Constructing Data Curation Profiles. *International Journal of Digital Curation* [on line]. 2009, vol. 4, no. 3, p. 93-103. <<http://ijdc.net/index.php/ijdc/article/view/137>> (consulted on 17 February 2016).
33. MIT Libraries. *Digital Content Management Infrastructure Improvement: FY13 strategic objective*. MIT Libraries, 2013.



34. Day, Michael; MacDonald, Ann; Kimura, Akiko; Pennock, Maureen. Implementing digital preservation strategy: developing content collection profiles at the British Library [on line]. In *Proceedings of JCDL '14: Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, London, UK, 8-12 September 2014 <<http://dl.acm.org/citation.cfm?id=2740773>> (consulted on 17 February 2016).
35. Pennock, Maureen; Wheatley, Paul; May, Peter. Sustainability assessments at the British Library: Formats, frameworks, & findings [on line]. In *iPRES 2014: Proceedings of the 11th International Conference on Preservation of Digital Objects 2014*. Melbourne, Australia, 6-10 October 2014 <[https://phaidra.univie.ac.at/detail\\_object/o:378110](https://phaidra.univie.ac.at/detail_object/o:378110)> (consulted on 17 February 2016).
36. Van der Knijff, Johan. *Assessing file format risks: searching for Bigfoot?* [on line]. Open Planets Foundation blog, 30 September 2013 <<http://openpreservation.org/blog/2013/09/30/assessing-file-format-risks-searching-bigfoot/>> (consulted on 17 February 2016).
37. Digital Preservation Coalition. *File formats assessments* [on line]. <[http://wiki.dpconline.org/index.php?title=File\\_Formats\\_Assessments](http://wiki.dpconline.org/index.php?title=File_Formats_Assessments)> (consulted on 17 February 2016).