

# Project to Production: Digital Preservation at the Houses of Parliament, 2010–2020

Christopher Fryer  
Houses of Parliament

## Abstract

The Parliamentary Archives is responsible for preserving and providing access to the historical records of the UK Parliament, a collection of national, and indeed international, importance which encompasses a wide range of digital content, created and used in an ever changing environment. Since 2010 a staged project has led to the successful implementation of an operational digital repository. The transition from project to production allows the Parliamentary Archives to reflect on the organisation's progress in digital preservation. However, the deployment of a production digital repository also allows the Parliamentary Archives to outline future goals. The project has demonstrated the viability of implementing digital preservation infrastructure. The challenge remains to embed digital preservation as a business as usual activity.

*Received* 16 January 2015 | *Accepted* 10 February 2015

Correspondence should be addressed to Christopher Fryer, Senior Digital Archivist, Parliamentary Archives, Houses of Parliament, London, SW1A 0PW, United Kingdom. Email: [fryerc@parliament.uk](mailto:fryerc@parliament.uk)

An earlier version of this paper was presented at the 10<sup>th</sup> International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



## Introduction

The Parliamentary Archives of the United Kingdom has recognised from an early stage the corporate need to ensure digital resources remain authentic and accessible over time. This paper describes how Parliament has implemented an operational digital repository, the rationale for doing so, and how we have addressed the various challenges and opportunities arising from this approach. Successful digital preservation entails much more than technological solutions. Cultural change and organisational development as well as technical challenges also play a key role in transitioning from a project to a viable service. We conclude that by forming part of a sustainable digital preservation service, the Parliamentary Archives fulfils an essential function of Parliament by securing access to exclusive digital resources for current and future generations.

## Discussion

### Background and Context

The Parliamentary Archives provides a records management and archives service for both Houses of the UK Parliament. It preserves and provides public access to an archival collection of international significance, including many of the core constitutional records of the UK. It also supports present-day business information management activities within Parliament. A major challenge for the Archives has been to develop the capability to preserve digital records alongside its traditional paper and parchment collections. Since 2010 a staged project has led to the successful implementation of an operational digital repository, which enables Parliament to preserve authentic born-digital records and digital surrogates, ensuring access for current and future generations, and mitigating threats such as cultural and technological change and the inherent fragility and mutability of digital information. Parliament's approach to digital preservation is described in detail in Parliamentary Archives strategy and policy.<sup>1</sup>

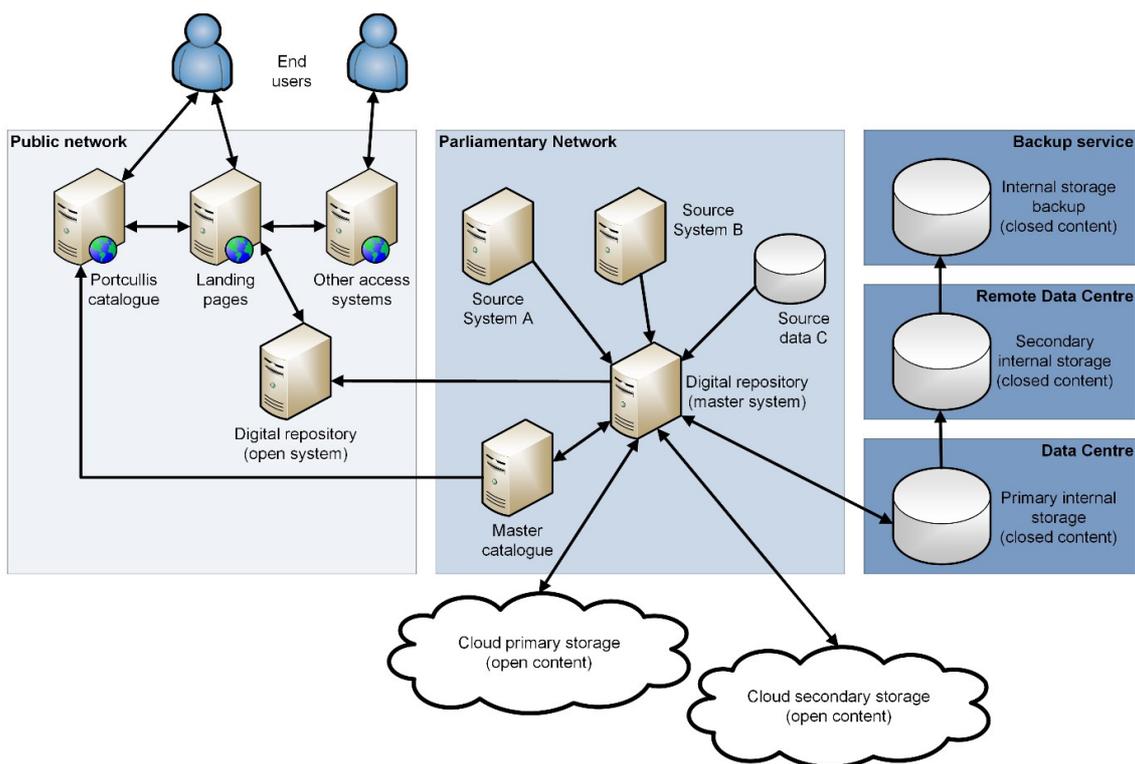
The Parliamentary Archives acquires growing volumes of parliamentary digital content from a broad range of sources, from born-digital business records managed in Parliament's Electronic Document and Records Management System (SPIRE) and the digital outputs of both Houses' substantial publishing activities, to digitised surrogates of iconic parliamentary documents. The Archives is also responsible for archiving the parliamentary web estate, which consists of the main parliament.uk domain, along with many other sites and third-party channels, such as social media. As of January 2015, the Archives had ingested over 14 TB of records into the digital repository, with at least 80 TB of priority content identified for ingest within the next four years. This volume will, of course, grow year by year. For example, with audio-visual content the collection could quickly grow to petabyte levels. Although the majority of its collections are open to the public, a proportion of material is closed; the digital repository must therefore also enforce access controls.

---

<sup>1</sup> Digital Preservation in Parliament: <http://www.parliament.uk/business/publications/parliamentary-archives/digitalpreservation/>

## Architecture

The heart of Parliament's digital repository is a commercial software platform (Preservica Enterprise Edition (formerly Safety Deposit Box) from Preservica, part of the Tessella Group<sup>2</sup>). The repository ingests content from a wide variety of internal systems and data sources. Descriptive metadata about digital objects is managed within the Archives' catalogue management system, alongside descriptions of the physical collections. The high level architecture of Parliament's digital repository is illustrated in Figure 1.



**Figure 1.** High level architecture of Parliament's digital repository.

Descriptive metadata is mirrored to the public, web-based front-end for the catalogue, called Portcullis<sup>3</sup>, and open content is replicated across to a second, public-facing instance of the repository. Public users can discover this content via the catalogue, from which they are directed to a landing page system that describes how the content can be accessed. In most cases, this provides direct access to view or download the object, but it may also link to an existing access system. In the case of copies which are not available online or where access is chargeable, users are directed to order copies.

Open content is stored in the cloud, with copies of all content mirrored between two different cloud service providers. The two cloud storage services operate on different technology platforms, one being based on EMC Atmos<sup>4</sup>, the other on Amazon S3 Web Services<sup>5</sup>. Closed content is stored on internal disk-based storage, mirrored between two data centres and with a traditional backup service.

<sup>2</sup> Preservica: <http://preservica.com/>

<sup>3</sup> Portcullis: <http://www.portcullis.parliament.uk/calmview/>

<sup>4</sup> EMC Atmos: <http://uk.emc.com/storage/atmos/atmos.htm>

<sup>5</sup> Amazon S3 Web Services: <http://aws.amazon.com/s3/>

Implementing the technical architecture for the digital repository has been a significant activity to undertake for Parliament in terms of resource and time. Established project management methodologies were used in order to carry out the implementation. The most encouraging aspect of this process was the collaboration between Archives staff and a much broader set of stakeholders. Ranging from Parliamentary ICT, to digital record producers across the business, the Digital Preservation Project demonstrated the viability of closely working with a wide variety of departments across a large and complex organisation. A number of factors contributed to this achievement including a robust business case, which documented in detail the justifications and objectives for digital preservation, along with management buy-in, which ensured appropriate high level representation across the business.

The main infrastructure challenges have arisen primarily from more generic issues around the nature of Parliament's network. For example, configuration of Preservica's services, which interact with elements of systems traditionally outside the internal network, have required fundamental changes to elements of Parliament's network infrastructure in order to comply with network policies. The other challenge was simply one of scale: it was essential to devise an architecture that would allow timely ingest of large data volumes, but without having an adverse impact on other systems within the Parliamentary network. This has required the use of dedicated servers and internet connections. However, although the technical challenges of integrating a digital repository with an existing organizational IT infrastructure should not be underestimated, they have all proven manageable.

## **Organisational Culture**

Experience of implementing a digital repository clearly demonstrates that digital preservation is not just a technological issue. It is the less tangible aspects of digital preservation that present the most challenging areas to overcome. Establishing any form of digital preservation capability revolves around the organisational context digital preservation efforts exist in. If we take the most obvious example of communicating highly conceptual digital preservation activities outlined in the Open Archival Information System (OAIS), the risk is that high-level policy framework governing the OAIS's activities soon become lost in communicating practical requirements. As Brian Lavoie states:

'A shared perception of these requirements serves as a point of familiarity in what can be an uncertain landscape; it is also a necessary condition for building well-understood, interoperable, and ultimately, trusted digital preservation systems' (Lavoie, 2014).

The fostering of shared perceptions of key requirements across internal and external stakeholders has been a critical factor in the implementation of digital preservation capability for the Houses of Parliament. For example, one benefit of provisioning the repository software to a commercial company has resulted in Archive staff efforts directed towards formalising processes to ingest priority at risk content.

Emphasis must be given to the fundamental approach which underpins Parliamentary Archive efforts to enact and react to organisational change as a result of digital preservation capability. A pragmatic, practical, and incremental approach has been undertaken over a significant period of time. Work began on drafting an initial business case in 2009, and in the same year a Digital Preservation Policy was finalised

and approved. The Digital Preservation strategy and roadmap were created to manage the risk to Parliament's digital resources up to 2013 and to create a sustainable model after that point. The business case itself detailed activities for 2010-15, taking the project through to business as usual. This timescale should be considered in light of the organisational context in which Parliament operates, and highlights the significant undertaking which introducing a new business process entails.

By examining this timeframe further, clear priorities and key points of interdependence begin to emerge. The business case required an initial high level effort to ensure that it would result in a positive outcome. It should not be underestimated the actual amount of work required to define requirements and develop them further into a fully formed and robust argument. Engaging with not just ICT, but also content creators proved vital in accurately assessing the current digital landscape without digital preservation capacity. A relatively straightforward Digital Asset Register identified priorities for action based on the vulnerability of the content, the proximity of the risk, and the cost of failing to take action greatly aided this effort.

## Workflows

The Parliamentary Archives working environment for digital preservation relies principally on its ability to anticipate and adapt to the changing needs of the organisation. A 2008 audit of Parliament's digital resources found that a small quantity of Parliamentary data which was of long-term value had already been lost and that an estimated 50 TB was in need of urgent attention. In practice this means that a workflow based system, which Preservica provides, enables practitioners to formally process digital content identified as part of a strategic effort.

A broad range of content has been identified; including business information and content from the Electronic Document Records Management System (EDRMS), digitised resources (e.g. Historic Hansard<sup>6</sup>), and web archive content<sup>7</sup>. The move towards improving the ways that digital business papers are provided has also resulted in an automated workflow, which retrieves files from Parliament's open data platform, extracts embedded files, and ingests them into the digital repository, synchronising the new content with the cataloguing system (CALM). By prioritising key areas identified as producing content which has long-term value, Parliament can ensure that access to digital resources is maintained, throughout their planned life cycle, preserving both active business information and information of permanent historical value for future users.

In order to develop workflows in the future it will be necessary to fully understand a variety of variables and responses identified. The overarching principle for the Parliamentary Archives is to render digital assets accessible to the relevant designated community. Carefully considered audit and certification procedures provide a means to provide a long-term trustworthy digital preservation capacity which is capable of sustaining this rendering ability. Although formal repository certification may be inappropriate for Parliament, the process that results from auditing infrastructure and procedures can play a significant role in benefits realisation. As demonstrated by the State and University Library of Denmark, self-assessment has led to improvements in specific tasks, optimising procedures – technically and organisationally – and a baseline to benchmark future improvements on (Elstrøm and Junge, 2014). A suitable self-

<sup>6</sup> Historic Hansard: <http://hansard.millbanksystems.com/>

<sup>7</sup> Parliament's Web Archive: <http://www.parliament.uk/business/publications/parliamentary-archives/web-archive/>

assessment and auditing framework is currently under consideration for the Parliamentary Archives digital preservation programme.

Many digital asset items are complex in nature and may consist of a number of files, each possibly of a different format. Consequently an evolution in our current ability to undertake preservation planning for digital assets is required. Contemporary work by the British Library on Sustainability Assessments exemplifies this approach:

‘This work suggests a new and more nuanced approach is necessary to avoid the comparative scoring of format against format and the focus on format obsolescence without consideration for more subtle and pressing preservation risks.’ (Pennock et al., 2014)

Ultimately, ensuring workflow functionality that results in digital assets which can continually be rendered can only be achieved through evidence based preservation planning.

### **Cloud Storage**

As part of its new digital repository infrastructure, the use of cloud storage services has enabled Parliament to provide a rigorous preservation storage capability that is flexible, scalable, and cost-effective. The decision to use the cloud in this case must be understood in the context of a wider drive to using cloud services within the UK public sector, and the development of the G-Cloud Framework.

Although not subject to this mandate, Parliament has also chosen to operate a ‘Cloud First’ policy. While it was decided that the digital repository management system should be managed in-house, the storage platform for the repository content was identified as a candidate for using the cloud. After a thorough options review, it was decided that open content would be stored with cloud storage providers, whereas sensitive closed content would be stored on an internal storage platform. Storage services were procured through the G-Cloud Framework, the first time this approach had been used by Parliament.

One key concern was to ensure Parliament’s ability to comply with relevant legislation, and in particular Freedom of Information and Data Protection laws, as well as addressing data sovereignty. In both cases, these concerns were addressed quite straightforwardly by ensuring that the relevant requirements were clearly defined and subsequently incorporated into contracts. It was a specific requirement that all data must be hosted entirely within the European Economic Area, to ensure that it is covered by European privacy legislation and not subject to other jurisdictions, and hence to legislation such as the US Patriot Act.

There are real risks that data might be lost if a supplier were to go out of business, in the event of a contractual dispute, or at the end of a contract. Instead of relying upon a single cloud storage provider, Parliament has procured two, which operate in parallel with all content duplicated between them. Each provider maintains multiple copies of all content, duplicated in at least two geographically-separate data centres, and uses techniques such as erasure coding to provide additional levels of durability. The two providers also operate on entirely distinct technologies, which offer a further degree of resilience and insulation from threats associated with a specific technology. The use of two suppliers increases the durability of Parliament’s storage (although quantifying this remains a future challenge), and provides insulation from supplier failure. The two

contracts have also been deliberately offset in time, to minimise the risks associated with any future changes of supplier.

Understanding the long-term economic implications of cloud versus traditional models remains a significant challenge. Parliament undertook a cost modelling exercise which indicated that, over an eight year period, the cloud would be significantly cheaper than in-house storage for the digital repository. However, this calculation inevitably included a number of assumptions regarding both internal and external factors, and its validity therefore rests upon their accuracy, which has yet to be fully tested.

Furthermore, the longer-term economics remain much more uncertain. Work by researchers such as David Rosenthal (2014) suggests that the cloud may prove a much more costly option for archival data storage over the long term, but further analysis is required in this area, alongside research into sustainable models for funding digital preservation over time – decisions about the economics of the long term data storage need to be based on a thorough understanding of how preservation activities will be funded over the same timescales.

Our experience thus far suggests that using the cloud, as Parliament has done, to provide one specific element of the digital repository infrastructure, is in many ways little different to providing that element in-house. However, it does tend to bring the risks and issues associated with digital repository storage in general to the fore, and has ensured that we have fully considered the risks and identified appropriate mitigations, both with respect to the cloud and our in-house infrastructure. This can only be a positive outcome.

Perhaps the most fundamental change necessitated by use of the cloud is the delegation of certain responsibilities to a third party, which requires issues of trust and transparency to be addressed. In particular, it requires customers to ensure that roles and responsibilities are defined with absolute clarity, so that there can be no doubt where the boundaries lie between the obligations and expectations of the customer and the supplier. Allied to this, clear, practical and appropriate service level agreements are essential.

It is essential to fully understand the associated risks and ensure that appropriate mitigations are in place. However, the risks arising from the cloud are often similar to those pertaining to other storage technologies, and strategies for managing them are available. In the longer term, questions about economics and portability remain, but these are not, in themselves, reasons to avoid using the cloud.

## **Collaboration**

Collaboration efforts played a vital contribution across all levels of activity. Parliamentary Archive collaborative initiatives involved working with internal stakeholders, but just as important was engagement and collaboration with external organisations that embarked on similar paths in building digital preservation services. Canvassing efforts undertaken during initial research resulted in a range of contacts, which ultimately still have relevance today. It is striking how willing organisations are to facilitate knowledge exchange across numerous efforts in a relatively small field of professionals.

These efforts resulted in a network of practitioners actively collaborating on common challenges which face their respective organisations. For instance, informal and formal digital repository user groups involving the Wellcome Trust, HSBC, British Library, the National Archives and many other European institutions now exist. These networks play a crucial role in reducing duplication of effort whilst also having the

added benefit of multiplying influence in leveraging software providers and communities. From the inception of the digital preservation profession collaborative efforts played a crucial role in progressing efforts. It is clear that in order for the digital preservation community to continue to progress collaborative efforts must evolve and deliver evidence based results.

Early efforts to incorporate a strategic digital preservation effort in Parliament relied upon the input of the wider community. In developing a policy and strategy which mandated the preservation and access of digital records, a thorough survey of other organisations was conducted. By engaging partners in developing a policy and strategy the Parliamentary Archives succeeded in benchmarking key requirements. This principle of engaging partners has endured and will continue to be utilised.

Current collaborative efforts are focused on developing preservation planning effectiveness. Current trials involve the Simple Property-Orientated Threat (SPOT) model, which aims to facilitate a typology of digital preservation threats which maximise conceptual clarity and avoid ambiguity and redundancy by clearly defining and organising threats in a simple and consistent manner (Caplan et al., 2012). Digital preservation processes are currently in a transition phase whereby they move from project based, to operational, and finally to business as usual. A key dependency in any business as usual activity is the ability to effectively plan, identify, and address present and future challenges. By engaging with institutions that are currently developing digital preservation planning capability the importance of effective collaboration shows no sign of decline.

## Vision

The Parliamentary Archives vision is clear:

‘Parliamentary records are at the heart of our democracy. They have embodied our liberties, rights and responsibilities for over five hundred years. We help Parliament work more efficiently and openly, enabling it to make its decisions and act as effectively as possible. And we want to inspire everyone with the compelling story of Parliament, people, and communities right up to the present day.’

Current digital preservation capacity plays an integral role in fulfilling this goal. It must be stressed that in order for digital preservation services to continue contributing towards this vision sustained progress is essential. This is not a simple case of increased resources. Present efforts are concentrated on embedding digital preservation within the Parliamentary Archives and more broadly across the organisation. The often overlooked clarity of language aspect is amplified when communicating with areas of the business outside the Archives department. It is increasingly common that the term digital preservation itself is only used when practitioners are sure the understanding is clear. Working in tandem with the Information and Records Management Service (IRMS), who are also part of the Parliamentary Archives, forms an integral element of contemporary and future digital preservation work. Framing combined records management and digital preservation goals as Digital Continuity provides clarity when engaging with the business<sup>8</sup>. If digital preservation efforts are to succeed in the coming

---

<sup>8</sup> The National Archives: Managing Digital Continuity: <http://www.nationalarchives.gov.uk/information-management/manage-information/policy-process/digital-continuity/>

years ensuring Digital Continuity provides a viable process to identify key touch points in Parliamentary processes is of paramount importance.

A clear theme which has emerged within the Parliamentary Archives is the professionalisation of contact with the wider organisation. This is particularly true of IRMS, who provide systematic control of Parliamentary business information and content (i.e. data, documents, records and other recorded information that has a specific content and value) throughout its life cycle so that the business can meet operational needs, legal requirements, and public expectations. Digital preservation now needs to meet these same levels of service in order to “help Parliament work more efficiently and openly, enabling it to make its decisions and act as effectively as possible”. Core services such as advice and consultancy, risk analysis, building capacity, and system support can be provided by working in conjunction with IRMS. Ultimately digital preservation, in collaboration with records management, can play a crucial corporate role by delivering a strategic digital continuity service for Parliament.

The corporate capacity to preserve digital records underlines a key current and future function for the Parliamentary Archives. Along with corporate responsibilities, providing public access to digital records is another fundamental area which has been identified for future development. Currently, records in all formats are provided to the public through the Archives online catalogue – Portcullis. There is little purpose to digital preservation efforts without making assets available to the appropriate designated community. To this end we are committed to enhancing current digital access arrangements in line with strategic objectives and appropriate resources. There is enormous potential to reach new audiences with engaging digital content, which has been made possible by digital preservation. If this potential is to be realised then significant progress is still needed in providing public access to digital records. This is a challenge which is faced by nearly all major memory institutions and the wider digital preservation community.

## Conclusion

The Digital Preservation Project has succeeded in delivering resilient, long-term access to vital digital content. By implementing digital preservation facilities Parliament has:

- Ensured that the long-term digital memory of Parliament is not lost or inaccessible, or compromised in any way which could damage either House’s ability to do its work, or its reputation;
- Protected Parliament’s investment in digitisation and corporate systems by securing digital assets of long-term importance, meeting Parliament’s legal and evidential needs, helping ensure continuity of service provision and enabling future re-use of digital content;
- Enabled Parliament’s mission to offer permanent public access to its online resources, for leisure, educational, academic or business use, and to support democratic accountability;
- Enabled Parliament to save future expenditure on the rescue or recreation of important digital assets of long-term value.

These strategic accomplishments are clearly crucial in determining the progress that Parliament has made over the last five plus years. However, they are only one indication of progress. Various lessons have been learnt which can be directly applied to future digital preservation work. For example, forging strong relationships with key stakeholders, especially such as ICT departments, ensures that stakeholders are aware of prospective strategic changes. Close collaboration with internal and external partners and the fruitful results this methodology produces are particularly encouraging for a complex organisation. By incrementally adopting pragmatic and practical goals the organisation has undertaken a steady transformation into one which is increasingly capable of ensuring digital continuity. Applying this approach is the direct result of understanding the organisational context in which the digital preservation service would operate.

Future challenges remain. Ingest of priority content continues and the embedding of business as usual procedures and policies is constantly maturing. Preservation planning is a relatively under-developed area of digital preservation. The Parliamentary Archives is currently focussed on progressing planning by embedding best practice risk analysis processes. Future developments in online access are also on the agenda which aim to dramatically improve the web presence and availability of the increasing number of digital content made available through digital preservation. As we turn to the next decade in digital preservation, the Parliamentary Archives can be confident in the assertion that further progress is achievable. Digital preservation is a continually evolving and challenging area which Parliament has taken crucial steps towards meeting. The aim for Parliament in the next decade is simple – to move forward by building on what has been learned.

## Acknowledgements

I would like to acknowledge the support of Adrian Brown, Director of the Parliamentary Archives, in writing this paper.

## References

- Caplan, P., Lavoie, B., & Vermaaten, S. (2012). Identifying threats to successful digital preservation: The SPOT model for risk assessment. *D-Lib Magazine*, 18(9/10). doi:10.1045/september2012-vermaaten
- Elstrøm, G. V., & Junge, J. G., (2014). Self-assessment of the digital repository at the State and University Library, Denmark: A case study. In S. Coates, R. King, S. Knight, C. A. Lee, P. McKinney, E. O'Meara, & D. Pearson (Eds.), *Proceedings of the 11th International Conference on Digital Preservation* (pp. 271–278). Retrieved from [https://www.nla.gov.au/sites/default/files/ipres2014-proceedings-version\\_1.pdf](https://www.nla.gov.au/sites/default/files/ipres2014-proceedings-version_1.pdf)
- Lavoie, B. (2014). *The Open Archival Information System (OAIS) reference model: Introductory guide* (2nd ed.; Technology Watch Report No. 14-02). Digital Preservation Coalition. doi:10.7207/twr14-02

Pennock, M., Wheatley, P., & May, P. (2014). Sustainability assessments at the British Library: Formats, frameworks and findings. In S. Coates, R. King, S. Knight, C. A. Lee, P. McKinney, E. O'Meara, & D. Pearson (Eds.), *Proceedings of the 11th International Conference on Digital Preservation* (pp. 141–148). Retrieved from [https://www.nla.gov.au/sites/default/files/ipres2014-proceedings-version\\_1.pdf](https://www.nla.gov.au/sites/default/files/ipres2014-proceedings-version_1.pdf)

Rosenthal, D. (2014, April 2). EverCloud workshop [Web log post]. Retrieved from <http://blog.dshr.org/2014/04/evercloud-workshop.html>