Towards a Collections Model for Preservation Planning at the British Library

|  |  |
| --- | --- |
| **Michael Day** | **Maureen Pennock** |
| *The British Library**London, United Kingdom**Michael.Day@bl.uk**https://orcid.org/0000-0002-1443-5334* | *The British Library**Boston Spa, Wetherby, United Kingdom**Maureen.Pennock@bl.uk**https://orcid.org/0000-0002-7521-**8536* |
|  |  |

**Abstract – The development of a framework for preservation planning at the British Library has highlighted the need for a more-structured understanding of its digital collections, in particular with regard to identifying the specific sets of objects that would be the focus of preservation plans. Work has recently commenced on developing a model of the Library’s collections to support this.**

**Keywords – Preservation planning, collection models**

**Conference Topics – Innovation**

1. Introduction

The British Library is currently in the process of developing a framework for preservation planning, e.g. through projects like the Integrated Preservation Suite (IPS) [1]. One of the things that has emerged from this activity is the need for a more-detailed understanding of the Library’s digital collections, especially with regard to identifying the specific collections (or object groups) that would need to be the focus of preservation plans. This poster presentation introduces an initial attempt to model the Library’s collections to support preservation planning and other digital preservation activities, including repository ingest and migration.

1. Collections as specific sets of objects

Becker, *et al*. [2] have distinguished between *preservation planning* at an abstract level and how the concept might be implemented in practice in the form of *preservation plans*. They defined the latter as, “specifying an *action plan* for preserving a specific set of objects for a given purpose.” They refer to that *specific set of objects* elsewhere as a “collection,” defining them at first in neutral terms as, “the set of digital objects or records for which a preservation plan is created,” adding that in technical terms, however, a collection would be “all of the objects that shall be treated with the same tool with identical parameter setting during the application of preservation actions.”

Preservation plans, therefore, can only really work when they are applied to sets of objects with (at least) some features in common, for example: a collection of eBooks in EPUB format.

1. The British Library context

The British Library has to date developed an understanding of its digital collections from two main directions. The first might be seen as a ‘bottom-up’ approach, focused on the practical needs of individual ingest streams or of specific collections or projects. The main problem with this approach is that it can be difficult to link these isolated collections and sub-collections into an integrated whole. The second is a ‘top-down’ approach based on the production of collection profiles for all of the Library’s major digital content types [3]. These profiles were designed to be a way for the digital preservation team to work with curators and collection owners across the Library to identify all relevant collections, to explore high-level digital preservation requirements, and to help specify preservation intent. The profiles are also reviewed on a periodic basis to ensure that they remain up-to-date and in-line with curatorial and user expectations.

In parallel, the IPS project has been designing and implementing a technical infrastructure based on a web-based ‘workbench,’ which in turn provides interfaces to: a ‘knowledge base’ of information about file formats and software, a repository for preserving software, and a facility for storing Library-specific preservation information, including policies, preservation plans, and collection profiles [1]. When the project was developing an initial template for preservation plans, it soon became clear that there was a need to be able to specify collections in a more specific way than that which was made possible by the Library’s collection profiles. The profiles do contain tables listing collections at a lower level, but these are typically not specific enough to form the basis of a preservation plan.

1. Modeling the library’s collections

In order to help identify the appropriate collection levels at which to apply preservation plans, therefore, the British Library is now attempting to produce a model of its digital collections.

There was relatively little prior work to base this upon, except for a few attempts to develop formal models and ontologies for digital library services [4, 5]. Perhaps more directly applicable were the metadata models and schemas developed for collection-level description (CLD) in the late 1990s [6]. These initiatives aimed to integrate heterogeneous collections into a single discovery framework and were based on a comprehensive analytical model of collections and catalogues developed by Heaney [7].

The Library’s initial ambitions were far more modest. The aim was to break down the high-level collection areas, e.g. as described in the collection profiles, and link them to collections at a lower level in diagrammatic form using the MS Visio tool (Fig. 1). These can then be used as a basis for further analysis.

This work has only just started, but the modeling so far has established at least four (approximate) layers of hierarchy. The top layer represents high-level collection areas (books, newspapers, sound content, etc.), which is then divided into ‘born-digital’ and ‘digitized’ categories (the Library’s collection profiles had expressly tried to integrate these, but the distinction immediately re-emerged when the model incorporated collections at a lower-level of granularity). Individual collections and sub-collections then feature in the lower levels, which will in turn need to be broken down further in order to identify those specific sets of objects that could be the focus of a preservation plan. This is the part of the modeling process that will merit the most attention in the future and which will determine the success of the approach.



Fig. 1: Web archive collections

1. Future work

The current focus of the collections modeling activity is pragmatic, intended to support IPS and to amplify the Library’s collection profiles. It will also provide input into a major project that is underway to migrate the Library’s digital collections to a new preservation repository system. The modeling of the Library’s collections is still very much a work-in-progress and it will be interesting to see where it might lead next.

REFERENCES

1. P. May, M. Pennock, and D. Russo, “The Integrated Preservation Suite: Scaled and automated preservation planning for highly diverse digital collections,” 16th International Conference on Digital Preservation, iPRES 2019, Amsterdam, The Netherlands, October 2019.
2. C. Becker, H. Kulovits, M. Guttenbrunner, S. Strodl, A. Rauber, and H. Hofman, “Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans,” *Int. J. Digit. Libr.*, vol. 10, pp. 133–157, 2009. DOI: 10.1007/s00799-009-0057-1.
3. M. Day, M. Pennock, A. Kimura, and A. MacDonald, “Identifying digital preservation requirements: Digital preservation strategy and collection profiling at the British Library,” 11th International Conference on Digital Preservation, iPRES 2014, Melbourne, Australia, October 2014.
4. L. Candela, D. Castelli, N. Ferro, Y. Ioannidis, G. Koutrika, C. Meghini, *et al*., *The DELOS digital library reference model: Foundations for digital libraries*, Pisa: ISTI-CNR, 2007.
5. M. A. Gonçalves, L. T. Watson, and E. A. Fox, “Towards a digital library theory: A formal digital library ontology,” *Int. J. Digit. Libr.*, vol. 8, pp. 91-114, 2008. DOI: 10.1007/s00799-008-0033-1
6. A. Powell, M. Heaney, and L. Dempsey, “RSLP Collection Description,” *D-Lib Magazine*, vol. 6, no. 9, September 2000. DOI: 10.1045/september2000-powell
7. M. Heaney, *An analytical model of collections and their catalogues*, 3rd issue, rev., Bath: UKOLN, January 2000. URL: http://www.ukoln.ac.uk/metadata/rslp/model/