The Integrated Preservation Suite

Demonstrating a scalable preservation planning toolset for diverse digital collections

Peter May
British Library, UK
peter.may@bl.uk
https://orcid.org/0000-0001-8625-9176

Maureen Pennock
British Library, UK
maureen.pennock@bl.uk
https://orcid.org/0000-0002-7521-8536

David Russo
British Library, UK
david.russo@bl.uk
https://orcid.org/0000-0003-2829-3936

Abstract – The Integrated Preservation Suite is an internally funded project at the British Library to develop automated and scalable preservation planning capability for a highly diverse and growing digital collection. Core components include a technical knowledge base, a software repository, a policy and planning repository, and a preservation watch function, all accessed via a web-based Preservation Workbench and designed to integrate with any repository software through a modular API-based architecture. We will demonstrate work to date, showing how preservation plans can be constructed and populated for the formats currently supported.

Keywords – digital preservation strategies, knowledge base, preservation planning, software preservation, preservation watch.

Conference Topics – The Cutting Edge – Technical Infrastructure & Implementation

I. Introduction

The Integrated Preservation Suite is an internally funded project at British Library that builds upon several years of preservation activities to develop and enhance the Library's preservation planning capability, largely focused on automation and addressing the risks and opportunities specific to the Library's heterogeneous digital collections. It aims to achieve this through the development and integration of several components – including a knowledge base, a software repository, a policy and planning repository, and a web-based workbench - designed to meet separate but complementary goals, all combined with the population of these components with content required for the infrastructure to work in a business environment.

This demonstration will showcase the functionality and contents of the Integrated Preservation Suite to date.

II. Overview

The Integrated Preservation Suite is intended to help us with format-based risk mitigation at scale and across all of our collections. We have developed the architecture and associated data models recursively through the course of the project, building components from the ground up to meet our needs where necessary. The project is a three-year initiative and we are currently in the final year of development, after which the suite will be deployed, maintained, and further populated as a business-as-usual concern.

Components reviewed in the demonstration will include:

Knowledge Base (KB): a graph-based curated knowledge base containing information from multiple different sources regarding formats, software, and wider technical environments relevant to the Library's digital collections;

- Preservation Software Repository (SR): a digital repository containing requisite current

16th International Conference on Digital Preservation
iPRES 2019, Amsterdam, The Netherlands.

and legacy software for rendering files stored in our digital repository and implementing preservation plans;
- Policy and Planning Repository (PPR): a document repository for storing collection-specific data including collection profiles, preservation policies, and collection-specific preservation plans;
- Preservation Workbench (PW): a web-based graphical user interface providing functionality for searching and curating the Knowledge Base, the Software Repository, and the Policy and Planning Repository, as well as for managing and developing format-specific preservation plans.

We will also discuss plans for the Preservation Watch - a component that monitors the archival store and our other IPS components in order to provide notifications to users regarding potential preservation risks, thereby instigating preservation planning.

Functions we expect to demonstrate will include:
- Preservation planning: Currently, our preservation planning approach is broadly following a SCAPE/Plato planning methodology, [1] bringing together various facets of information about a collection at risk to define the plan requirements, evaluate different strategies to mitigate any risks, analyze the results, make a recommendation, and construct an executable plan.
- Curating incoming data for the Knowledge Base: Data in the Knowledge Base is aggregated from several different sources; the Workbench provides a web-based interface to allow an appropriate user to compare incoming data with existing data and make decisions about how to proceed with each datum.
- Searching the Knowledge Base: Keyword searching is supported, e.g., a user can search for "PDF" or "Adobe" or any other term. This matches on key properties within the Knowledge Base, such as the file format/software name or an extension. To facilitate more in-depth queries, such as for identifying software that can migrate file formats, we provide a set of search labels with which to tailor queries.

Information on the overall architecture, data models, and development process will also be available. Staff will be on hand to discuss these and non-technical elements of the project including licensing challenges (and progress) for content in the software repository.

III. Required infrastructure and/or resources

We will need to use our own laptops for the demo. Wifi will be required. A projector may also be required depending on the room set up.

The duration of the demo can be tailored depending on how long you make available. We are not planning a structured session.

Goals of the demo:
- Showcase the integration element of the Suite as a pre-requisite for scalable preservation planning across diverse collections
- Engage with attendees to gather feedback that might inform future development or maintenance work

Acknowledgment

We would like to thank the other IPS team members whose dedicated work has helped shape and deliver the project: Michael Whitmore, Kevin Davies, Michael Day, Akiko Kimura, Simon Whibley, and John Beaman. We would also like to thank the National Library of Australia for kindly supplying us with a spreadsheet of their collected data, and the maintainers of File-Extensions.org for giving us permission to scrape their website.

References

[1] M. Kraxner, M. Plangg, K. Duretec, C. Becker, and L. Faria, "The SCAPE planning and watch suite," in Proceedings of the 6th International Conference on Digital Preservation, iPRES 2013., Lisbon, Portugal, September 2 – 6th, 2013 [Online]. Available: http://hdl.handle.net/11353/10.378091.