planets
A Research Infrastructure for Preserving Digital Objects

Rainer Schmidt, Ross King, Andrew N. Jackson

INTRODUCTION

Preserving the enormous volumes of data that are being produced and stored in data repositories and digital libraries is an important challenge. As digital information highly depends on a rapidly changing hard and software environment, data requires curation in order to keep pace with the technical evolution. In addition to the physical preservation of the content bit-streams, one must ensure the long-term interpretability in order to prevent a loss of information. A major requirement in this domain is the development of it-infrastructures that can automate the required preservation processes.

The project Permanent Long-term Access through NETworked Services (PLANETS) focuses on preservation requirements faced by cultural heritage institutions. The PLANETS research infrastructure provides an integrated environment for the development and evaluation of logical preservation strategies. Key research goals are the integration of diverse repository systems, preservation services, and the automated evaluation of preservation results.

RESEARCH GOALS

- ❑ Ontology for Formats and Characteristics
- ❑ Languages and Tools for Extraction and Comparison of Object Characteristics
- ❑ Vocabulary for Expressing Preservation Actions and Outcomes
- ❑ Integration of Diverse and Distributed Digital Repositories
- ❑ Unified Data Model for Digital Objects
- ❑ Transformations from/to different Digital Object Representations
- ❑ Distributed Access to a Large Variety of Preservation Tools
- ❑ Workflow Definition and High-Level Composition
- ❑ Recording of Provenance and Preservation Metadata (PREMIS-based)
- ❑ Deposit of Experiment Results and Data

APPROACH

Map and Register Data:
In order to make a digital collection available for experimentation, it must be first registered with the Planets data registry. This involves the mapping of the imported data items to the Planets Digital Object Model. Digital Objects encapsulate the concept of single digital entities and may be composed of one or more byte-streams and be associated with metadata. Planets Digital objects are stored and made available through a metadata repository.

Execution and Deposit:
Preservation workflows are executed against a digital collection using the Planets workflow engine. During execution, provenance and other metadata is automatically recorded by the workflow engine. A workflow execution is reflected and expressed through the digital object model, and stored through the data registry service. Workflow results can be directly deposited to a repository system or be manually accessed from the Planets data repository.

Data Curators
Repositories
Data Models
Encodings

Continuous Preservation Cycle
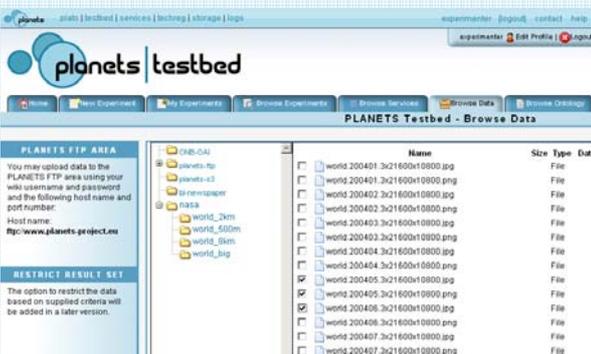
Metadata
Tools & Platforms
Services
Workflows

Develop a Preservation Strategy:
The Planets environment provides service-based access to a large range of preservation tools. Planets Preservation services define atomic digital preservation verbs, for example: *Migrate*, Characterize, Validate or Compare. The services provide access to different tools deployed in variety of environments. A graphical decision support environment as well as the Planets Testbed Application aid users in creating and evaluating specific preservation strategies.

Choose and Configure Experiment:
Once an executable preservation plan has been created for a specific scenario, it can be made available to other users by uploading it as a template to workflow repository. Workflow templates are build from a set of basic building blocks and specify an abstract execution process based on preservation components and decision logic. Users can easily reuse and individually configure preservation an existing workflow template based on generated XML workflow configuration files.
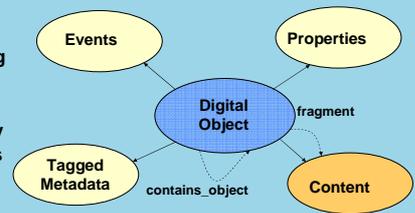
ACCES VIA PORTAL APPLICATIONS

The Planets Testbed Application [2]

SUMMARY

The Planets Framework [3] provides a technical infrastructure for accessing a broad range of preservation tools and data repositories. Interoperability is achieved by enforcing defined preservation operation and service profiles, based on a service-oriented architecture. The infrastructure is utilised for experiments by Planets applications through a common set of services. These include support for user management, data and metadata handling, service discovery, as well as workflow execution and monitoring. The project provides a public deployment* as well as a build-infrastructure and a formal process for testing, deploying and registering new preservation services.

* http://testbed.planets-project.eu/testbed/

DIGITAL OBJECT MODEL

- ❑ A minimal data abstraction that can be mapped against records retrieved from existing repository systems.
- ❑ Represents entities that are consumed and/or produced by the services within the Planets infrastructure.
- ❑ Encapsulates provenance and preservation metadata.
- ❑ Individual Digital Object Managers exhibit a common interface for accessing a number of institutional repositories.

REFERENCES

[1] R. Schmidt, R. King , A. Jackson, C. Wilson F. Steeg and P. Melms. A Framework for Distributed Preservation Workflows, In Proceedings of iPres 2009, 5-6 October 2009, San Francisco, USA and the IJDC (to appear).
[2] B. Aitken, P. Helwig, A. N. Jackson, A. Lindley, E. Nicchiarelli, and S. Ross. The Planets Testbed: Science for Digital Preservation. In Code4Lib Journal, volume 1(5), 2008.
[3] R. King et al. The Planets Interoperability Framework – An Infrastructure for Digital Preservation Actions. In Proceedings of the ECDL 2009, Cofu, Greece.

ACKNOWLEDGEMENTS