The Planets Interoperability Framework*

An Infrastructure for Digital Preservation Actions

Ross King[1], Rainer Schmidt[1],
Andrew N. Jackson[2], Carl Wilson[2], and Fabian Steeg[3]

[1] Austrian Research Centers GmbH - ARC
Donau-City-Strasse 1
1220 Vienna, Austria
firstname.lastname@arcs.ac.at
[2] The British Library
Boston Spa, Wetherby, West Yorkshire
LS23 7BQ, United Kingdom
firstname.lastname@bl.uk
[3] Universität zu Köln
Albertus-Magnus-Platz
50923 Cologne, Germany
fabian.steeg@uni-koeln.de

Abstract. We report on the implementation of a software infrastructure for preservation actions, carried out in the context of the European Integrated Project *Planets* – the Planets *Interoperability Framework* (IF). The design of the framework was driven by the requirements of logical preservation in the domain of libraries and archives. The IF is a Java-based software suite built on a number of open source components and Java standards. Specific features of interest are a web service architecture including specified preservation service interfaces for the integration of new and existing preservation tools and a workflow engine for the execution of service-based preservation plans.

1 Introduction

We describe a software infrastructure developed in the context of the EU Integrated Project Planets known as the Planets Interoperability Framework (IF). A more detailed overview of the Planets project can be found elsewhere [1].

2 State of the Art

Significant effort in the preservation community has been dedicated towards producing OAIS-compliant archiving infrastructures, although OAIS itself is primarily a conceptual model and does not provide guidelines for implementation [2].

* Work presented in this paper is partially supported by the European Community under the Information Society Technologies (IST) Programme of the 6th FP for RTD - Project IST-033789.

M. Agosti et al. (Eds.): ECDL 2009, LNCS 5714, pp. 425–428, 2009.

While Planets has concentrated on the specific OAIS component "Preservation Planning", the CASPAR project [3] places an emphasis on representation information and a complete OAIS-compliant archive implementation.

One component of the overall CASPAR architecture that is comparable to the Planets Interoperability Framework is the Preservation Data Store (PDS) [4]. The PDS approach delegates preservation-related functionality to the storage component. As in Planets, the PDS architecture supports integration with existing archives. However, PDS lacks the flexibility of the Planets IF plug-in approach to preservation tools and workflows.

The closest parallels to the Planets approach are the Australian project PANIC (Preservation services Architecture for New media and Interactive Collections) [5] and The United Kingdom's National Archives Seamless Flow framework [6]. Both are based on service-oriented architectures (SOA) in which preservation actions are invoked through Web Services.

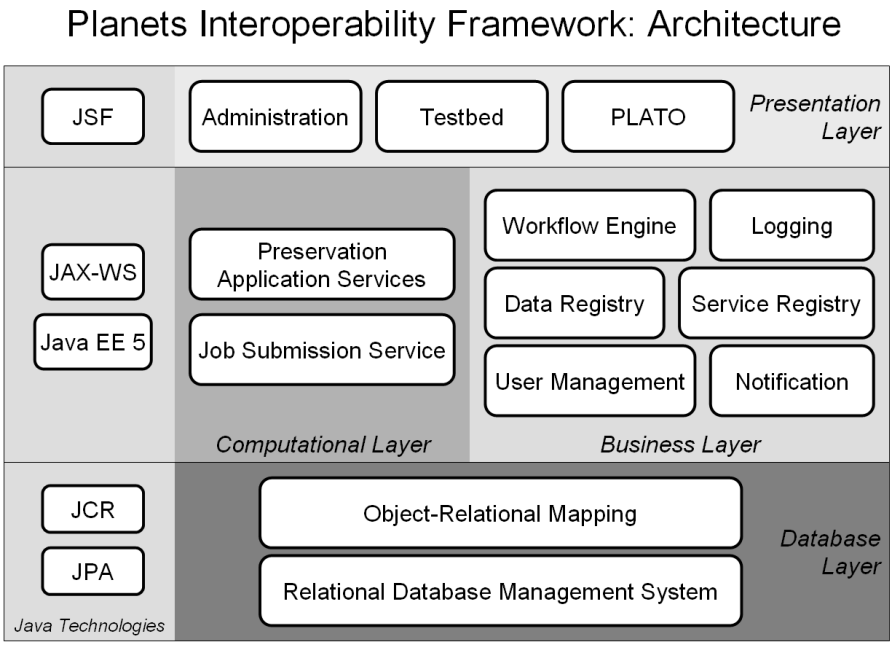

Fig. 1. Planets Interoperability Framework Architecture

3 The Planets Interoperability Framework

The Interoperability Framework provides an infrastructure to carry out digital preservation actions in the form of flexible, service-based workflows. Planets is driven by the requirements of memory institutions, primarily national libraries and archives. These institutions generally already have archiving systems in place, and replacing such systems is neither feasible nor desirable. Therefore

the IF was designed to run in parallel with existing archive systems; it is not intended to replace these or to provide archiving functionality.

For this short description of a selection of IF components and their functionality, refer to the *Business Layer* of figure 1. The *Data Registry* provides storage and persistence services to IF users, components and services. In particular the storage API provides methods for storing complex Digital Objects and Preservation Events. The *Service Registry* enables users and service providers to look up and publish information about preservation services, and enables Planets system administrators to manage information about these services. In order to support service discovery, we provide extensible service categorization mechanisms that allow the Service Registry to be queried using these categories. The Planets IF *Workflow Engine* implements a component-oriented enactor that governs the orchestration of the preservation components, including functionalities like session-management, communication, and preservation metadata handling. Distributed preservation workflows are conducted from high-level components that abstract the underlying protocol layers.

3.1 Implementation Details

The Interoperability Framework provides a Java-based infrastructure that leverages a number of standards and open source tools. Referring to the *Java Technologies* indicated in figure 1, the core of the IF implementation is the Java Platform, Enterprise Edition (Java EE 5) standard, which among other things provides a framework for the efficient implementation of Web Services and Web applications. In particular we make use of Sun Microsystems' Web Services Interoperability Technology (WSIT) suite and the underlying JAX-WS (Java API for XML Web Services) standard. The IF provides a pre-configured JBoss[1] application server as its default deployment environment. We chose JBoss as a stable, well-supported open source implementation of the Java EE 5 standard. Data persistence is provided through the Java Persistence API (JPA), supported by the underlying Apache Derby[2] relational database management system (RDBMS). Derby was chosen because its small footprint and pure Java implementation allow it to be easily packaged and installed with minimum user expertise. However, as we use only the standard Java Database Connectivity (JDBC) API and no RDBMS-specific features, it is possible to configure the IF for operation with other open-source or commercial databases in a production environment.

3.2 Performance and Scalability

A crucial aspect of our preservation system is the establishment of a distributed, reliable, and scalable computational tier. Advances in virtualization allow the deployment of entire preservation environments, including operating systems and applications, to distributed computational nodes. This allows one to instantiate sets of transient system images on demand, which can be federated as a

[1] `http://www.jboss.org/`
[2] `http://db.apache.org/derby/`

virtualized cluster. We have implemented a prototype Job Submission Service (JSS) that can manage such infrastructures and execute Planets preservation workflows on a virtual computing cluster or *Cloud*. Initial experiments with the Amazon Elastic Compute Cloud (EC2) reported elsewhere [7] have demonstrated the feasibility of this approach for the Planets service architecture.

4 Conclusions and Future Work

After the third project year, the Planets Interoperability Framework provides a stable preservation infrastructure, which is available for download[3] in a platform-independent installation package. In the fourth and final year of the project, we will continue research on scalable services for preservation actions based on virtual computational clusters. In addition, field tests of the Planets Software Suite will be carried out at partner institutions, demonstrating how the Interoperability Framework and the associated Planets applications and services can act as an added-value preservation action system for existing digital repositories at national libraries and archives.

References

1. Farquhar, A., Hockx-Yu, H.: Planets: Integrated Services for Digital Preservation. International Journal of Digital Curation 2(2) (2007)
2. ISO Standard 14721:2003: Space Data and Information Transfer Systems – A Reference Model for an Open Archival Information System (OAIS). International Organization for Standardization (2003)
3. Giaretta, D.: The CASPAR Approach to Digital Preservation. The International Journal of Digital Curation 2(1) (2007), http://www.ijdc.net/ijdc/article/view/29/32
4. Factor, M., Naor, D., Rabinovici-Cohen, S., Ramati, L., Reshef, P., Ronen, S., Satran, J., Giaretta, D.: Preservation DataStores: New storage paradigm for preservation environments. IBM Journal of Research and Development on Storage Technologies and Systems 52(4/5) (2008)
5. Hunter, J., Choudhury, S.: A semi-automated digital preservation system based on semantic web services. In: JCDL 2004: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, pp. 269–278 (2004)
6. Brown, A.: Developing Practical Approaches to Active Preservation. The International Journal of Digital Curation 2(1) (2007), http://www.ijdc.net/ijdc/article/view/37/42
7. Schmidt, R., Sadilek, C., King, R.: A Service for Data-Intensive Computations on Virtual Clusters. In: INTENSIVE 2009: Proceedings of The First International Conference on Intensive Applications and Services (to appear)

[3] http://gforge.planets-project.eu/gf/project/if_sp/