Building a Trusted Distributed Archival Preservation Service with iRODS

Jewel H. Ward, Terrell G. Russell, and Alexandra Chassanoff
School of Information and Library Science, University of North Carolina at Chapel Hill

Abstract

The Distributed Custodial Archival Preservation Environments (DCAPE) [1] project is a proof-of-concept of the viability of a distributed custodial preservation approach as a production service using iRODS [2]. The DCAPE partnership is comprised of 32 people across 10 institutions, as well as doctoral and graduate students. These partners include technologists, librarians, and archivists from computer science research, state libraries, and state, university, and cultural archives. In the first phase of this project, the project team created an initial set of 26 iRODS rules for use by the DCAPE archivists and administrators. We based these rules on the stated needs of the partner organizations and the standards developed to determine trustworthiness by the ISO Mission Operations Information Management System repository assessment criteria (ISO MOIMS-RAC) group [3,4]. Within the DCAPE project, the iRODS infrastructure serves two critical functions: (1) it validates the trustworthiness of the repository through the enactment of ISO-MOIMS-compliant policies; and, (2) it enables the distributed auditable administration of the repository through the invocation of iRODS rules.

Index Keyword Terms— Distributed Systems, Document/file management, Data Sharing, Online Information Services

1. Introduction

iRODS is extensible; rules may be added or deleted as needed by archivists and administrators. These rules are machine-actionable policies; that is, a repository administrator may enforce written policies at the machine level by re-writing the policies as code that can be read by the iRODS rule engine. These policies may also be enforced across a distributed environment, which provides a method for archivists in disparate locations to manage their records. The rule engine is dynamic; the rules may be installed or removed as the needs of the repository administrators change.

In the first phase of the 2-½ year DCAPE project, the team developed an initial set of 26 capabilities, based on the standards developed by the ISO MOIMS-RAC group. These capabilities include rules that manage Open Archival Information Systems (OAIS) compatible SIPs, AIPs, DIPs, identifiers, audit and security information, and enforce policies and service level agreements (SLAs), among others. The goal is to converge towards a set of fundamental building blocks necessary to create a distributed, trustworthy archival production system.

Preservation-as-a-Service (PaaS) is a potential business model that may prove viable as the technologies involved have become commodities and the availability of significant amounts of storage is common. Alongside bandwidth costs, the overhead associated with the management of preservation policy has become a limiting factor when dealing with large datasets. DCAPE aims to reduce this overhead and provide an archivist-level interface into a very powerful system.

In this poster, we describe the rationale for creating the DCAPE project, what we have accomplished so far, what we have learned, and what our future work on the project will entail.

2. Acknowledgements

This work is funded by the NHPRC Records Projects grant NHPRC RE10010-08, “Distributed Custodial Archival Preservation Environments” (2008-2011). We would like to think Richard Marciano for his comments on this poster.

3. References

[1] DCAPE project (Distributed Custodial Archival Preservation Environments), http://dcape.org.
[2] DICE, "Overview: Data Management & iRODS", iRODS Publications, San Diego, CA, March 19-20, 2009.
[3] MOIMS-RAC: Repository Audit and Certification working group, http://wiki.digitalrepositoryauditandcertification.org.
[4] M. Day. "Toward Distributed Infrastructures for Digital Preservation: The Roles of Collaboration and Trust", Journal of Digital Information, 1(3), College Station, TX, 2008.