Global Media Curation Platform Using iRODS

Yuma Matsui, Qian Liu*, Nathan Brock*, Jeffrey D. Weekely**, Dana Plepys****

*California Institute for Telecommunications and Information Technology (Calit2), UC San Diego
**MOVES Institute, Naval Postgraduate School
***Electronic Visualization Laboratory (EVL), Department of Computer Science, University of Illinois at Chicago

Abstract

iRODS data grid middleware is used to manage research data [1], and we also use it to curate and preserve super-high-definition media content including 4K movies in CineGrid Exchange [2]. We are building a media curation platform that enables long-term data preservation and streamlined workflows for curators to handle large amount of media data in a cross-organizational environment. The platform consists of a front-end web catalog system and a back-end storage and workflow system. The front-end is a web-based content management system and brings ease-of-use in metadata management and data retrieval. The back-end is an iRODS-based distributed storage system that guarantees data consistency and redundancy. It also provides an event-based rule engine to realize flexible workflows.

This poster illustrates the workflows and implementation of our media curation platform. We have geographically distributed iRODS servers with rules for our established workflows. We are testing the system and going to expand it to a global scale.

Index Keyword Terms— *Curation Workflow, Data Management, Data Preservation, Distributed Storage*

1. Workflow and implementation

The main workflows of curators in CineGrid Exchange are media data submission and media data access. In media data submission, curators first receive content from creators and evaluate its quality and validity. Then they collect metadata about the content and put it into the system through the web catalog interface. These metadata include category (Animation, Live Action, and Still Images), subject (Art, Entertainment, Scientific, Sports, and so forth), year, contributor, title, and other media properties. Metadata values are transferred from the front-end to the back-end in a predefined XML protocol. The back-end has an iRODS rule to read the received metadata values and create a human-readable directory structure that serves as a content upload destination for curators. Finally, the back-end system notifies curators of the content placeholder path, and curators upload media data with efficient iRODS client programs such as iDrop. The back-end is also equipped with an iRODS rule to replicate the uploaded files to other geographically dispersed iRODS servers.

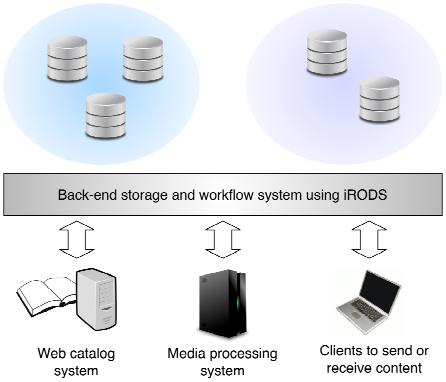

Figure 1. System architecture of CineGrid exchange.

In media data access, curators can first browse or search media data on the web catalog, and it returns paths of content in the back-end system. Then they can directly access iRODS to retrieve content with client programs.

Fig. 1 shows the general system architecture of CineGrid Exchange. The back-end consists of federated iRODS servers that periodically try to detect file corruption and node failure. The web catalog and the back-end communicate by exchanging messages. We will also integrate a high-performance media processing system with the back-end storage system using a messaging mechanism in the future.

2. References

[1] M. Hedges, A. Hasan, and T. Blanke, “Curation and preservation of research data in an iRODS data grid,” in e-Science and Grid Computing, IEEE International Conference on, 2008, pp. 457–464.

[2] S. Liu et al., “CineGrid Exchange: A workflow-based peta-scale distributed storage platform on a high-speed network,” Future Generation Computer Systems, vol. 27, no. 7, pp. 966-976, Jul. 2011.