

Rule-based Data Management Systems

Reagan Moore

San Diego Supercomputing Center (SDSC), USA

moore@sdsc.edu

Shared collections are being assembled on international scales. The shared collections support not only collaborative research, but also publication of data and preservation of data. Typical sizes of the shared collections are measured in Terabytes to hundreds of Terabytes, with collection containing a million files to tens of millions of files. We observe that the labor required to manage the collections scales with the collection size. We also observe that many of the tasks related to authenticity and integrity validation can be automated. The iRODS - integrated Rule-Oriented Data System is data grid software that automates the application of management policies, greatly reducing the effort needed to manage internationally shared data. A demonstration will be given of the software, along with sample rules for controlling the execution of remote micro-services that implement data grid operations.