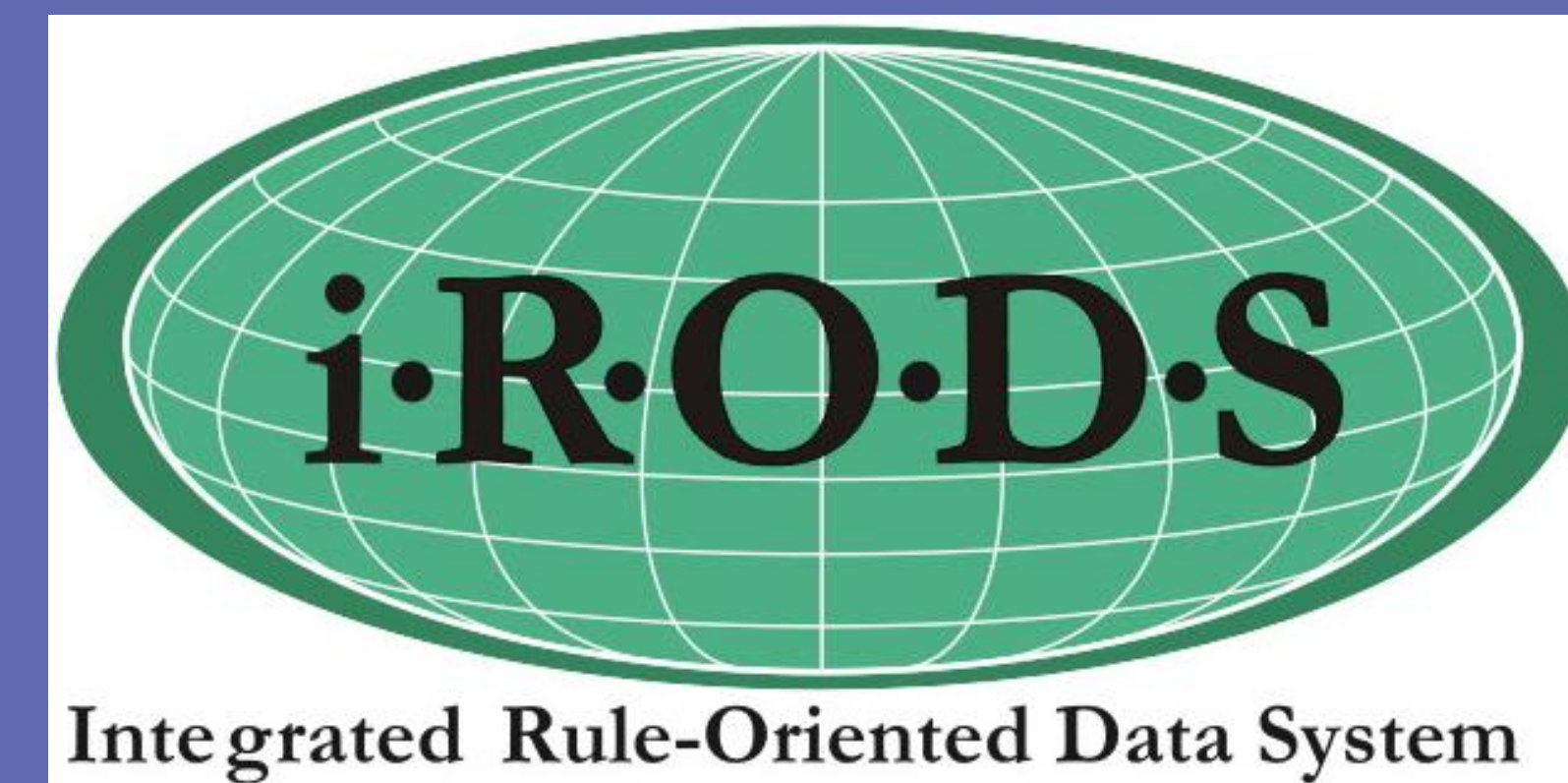




Implementing iRODS for Data Federation

Tiffany Mathews, Brandi Quam, Andrei Vakhnin,
& Michael Little

NASA Langley Research Center, Hampton, VA



ASDC Introduction

The Atmospheric Science Data Center (ASDC) at NASA Langley Research Center is responsible for the ingest, archive, and distribution of NASA Earth Science data in the areas of radiation budget, clouds, aerosols, and tropospheric chemistry. The ASDC specializes in atmospheric data that is important to understanding the causes and processes of global climate change and the consequences of human activities on the climate. The ASDC currently supports more than 44 projects and has over 1,700 archived data sets, which increase daily. ASDC customers include scientists, researchers, federal, state, and local governments, academia, industry, and application users, the remote sensing community, and the general public.

Strategy & Innovation

In 2012, the ASDC facilitated the development of the first ever strategic plan intended for fiscal year 2013 and beyond. The 2013 Strategic Plan serves as a mission-focused plan with six defined goals, each with supporting objectives and tasks for implementation that emphasize the vision and support the mission and values of the ASDC.

Goal #1

The ASDC will strive to expand beyond its existing customer base by increasing accessibility to a broader, worldwide market; through the use of innovative technologies, the ASDC will enhance data access capabilities and develop plans to share data with new user communities.

Pursuant to this goal, the ASDC is exploring and piloting new technologies to implement for enhanced data access capabilities. Through the implementation of iRODS, ASDC intends to expand the discovery and access of data holdings. This will be made possible as iRODS provides the ASDC's Data Products Online (DPO) an ability to federate with other NASA centers including the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center (GSFC). This federation and sharing of information will enable the ASDC and NCCS to utilize multi-year and multi-instrument data as well as improve and automate the discovery of heterogeneous data, increase data transfer latency, and meet customizable criteria based on data content, data quality, metadata, and production.

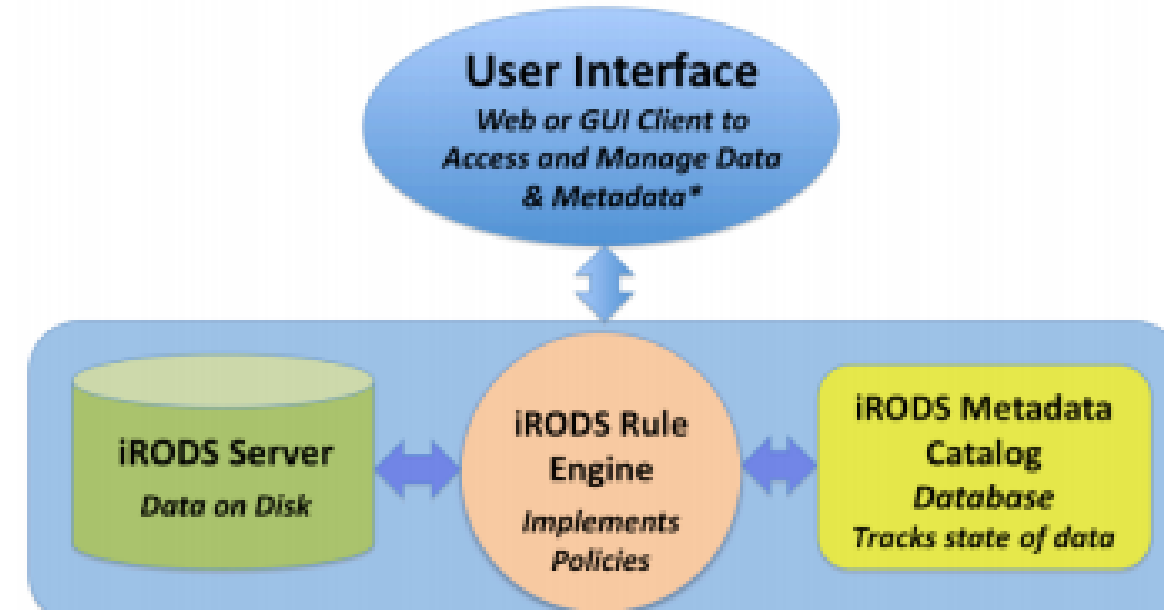
What is iRODS?

iRODS: Integrated Rule-Oriented Data System

iRODS is a Data grid software system developed by the Data Intensive Cyber Environments (DICE) group, (developers of the SRB, the Storage Resource Broker), and collaborators.

Data grids enable parallel downloads of datasets from selected replica servers that can be located in different locations but still accessible by users worldwide.

iRODS Data System Components



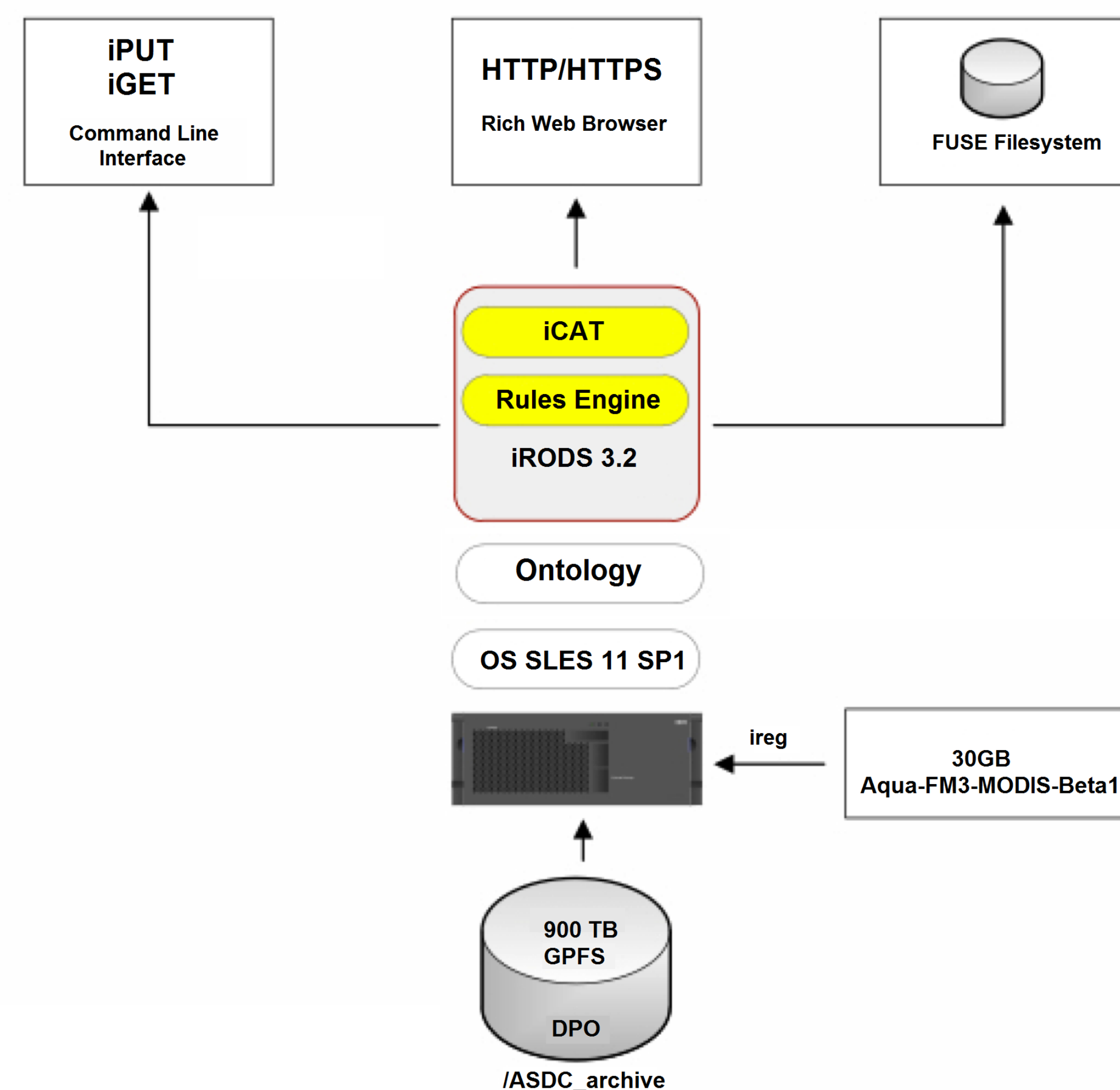
iRODS interfaces (GUI, Web, WebDAV, command line) let users search for, access and view, add/extract metadata, annotate, analyze and process, manage, replicate, copy, share, repurpose, control and track access, subscribe, and more.

iRODS Server software and Rule Engine run on each data server. The iRODS iCAT Metadata Catalog uses a database to track metadata describing data and everything that happens to it

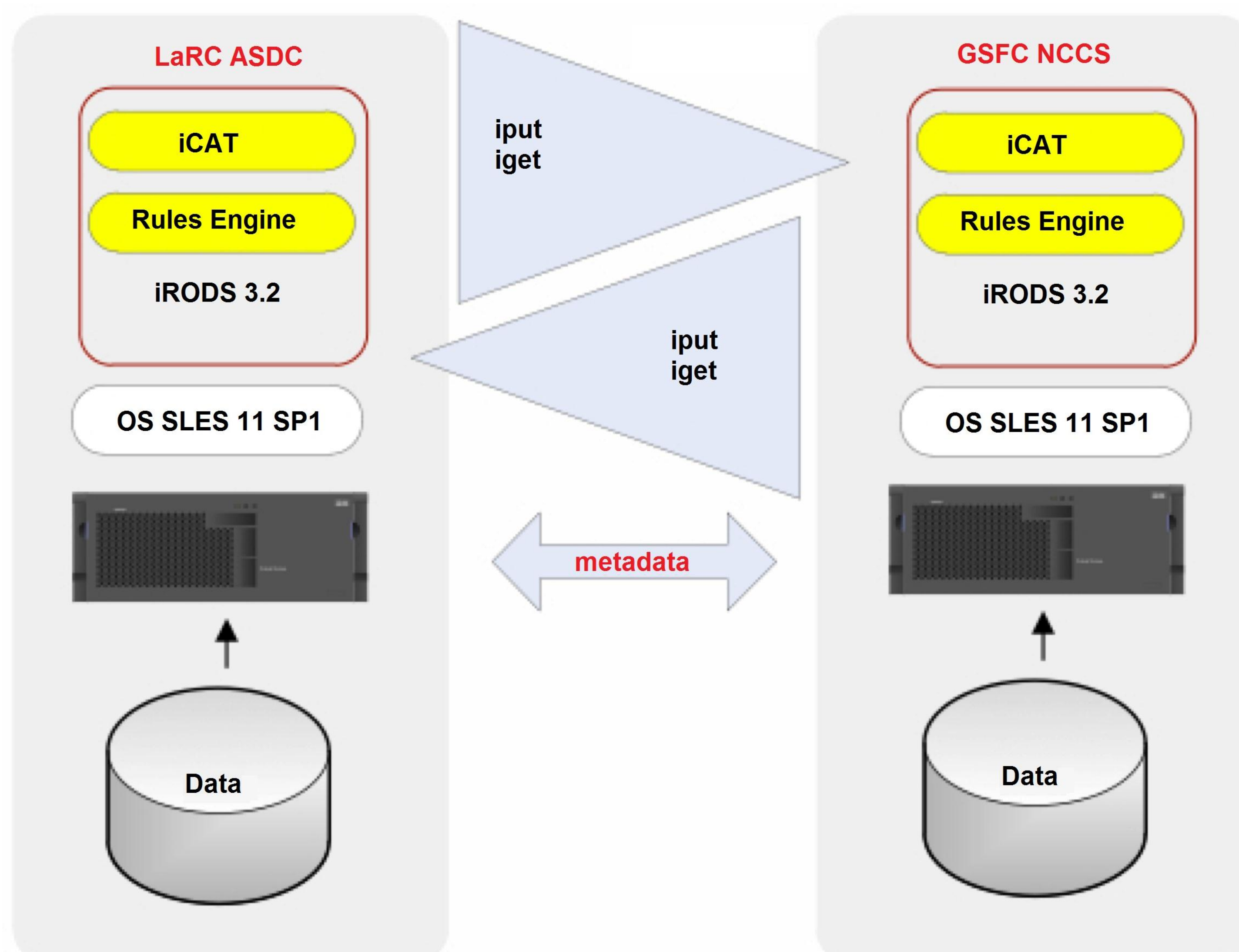
ASDC iRODS Implementation

The Atmospheric Science Data Center's implementation mission with iRODS is to establish a consolidated, shared services capability that will provide higher quality, more cost effective and efficient services, and greater access to our current science community users and future diverse customers.

ASDC iRODS Configuration



iRODS Federation - Flexible Architectures



-iRODS Federation Between the ASDC and NCCS

Way Forward

The iRODS data grid is highly extensible and can be used for data-intensive distributed computing – it is capable of managing both external and internal technology evolution.

The overarching goals of the ASDC in the implementation of iRODS are to:

- Maintain RENCI partnership to ensure seamless transition as new capabilities emerge; obtain e-iRODS membership (Summer 2013) in order to take advantage of extensive iRODS support services
- Operationalize iRODS pilot and leverage the iRODS architecture to extend capability for multi-DAAC (Distributed Active Archive Center) federation and distributed search
- Develop an approach to enable virtualization and provide capacity to respond in an agile way to new customer requests
- Provide a path to migrate existing services into the cloud and integrate cloud storage with our repository
- Integrate data discovery, management, and access applications (OPeNDAP, Hadoop, etc.)
- Preserve the integrity, credibility, and security of ASDC data holdings by leveraging micro-services and policy-based data management features of e-iRODS.

The ASDC will implement iRODS through three successive phases:

Phase 1:

- Establish collaborations with NCCS and RENCI;
- Test local implementation and federation of iRODS with NCCS (ETA June 2013);
- Demonstrate test results in terms of data transfer rates in order to address how the ASDC will collect consumption statistics with iRODS to feed into the EMS (ESDIS (Earth Science Data and Information System) Metrics System); and
- Identify steps and tools necessary to feed Precision Ontology into iRODS (AllegroGraph).

Phase 2:

- Fully implement iRODS at the ASDC, including feeding metadata from ontology (ETA October 2013);
- Operationalize iRODS; as well as
- Expand the collaboration with RENCI and NCCS and identify other collaborations.

Phase 3:

- Evaluate the feasibility of data access from the Amazon Web Services (AWS) cloud using iRODS (ETA December 2013); and
- Leverage iRODS to channel experiments into the cloud computing environment with ASDC data products.

Acknowledgements

The authors would like to thank Reagan Moore, Charles Schmidt, and Arcot Rajasekar at RENCI for their continuous support and collaboration with our iRODS implementation. Additionally, the partnership with NCCS (Daniel Duffy, John Schnase, Al Settell, Glen Tamkin, Ed Luczak, and Mark McInerney) has been invaluable to the success of this pilot.

References

"FACT SHEET: iRODS integrated Rule Oriented Data System" [Data Intensive Cyber Environments Center \(DICE\). Web](#)