



NASA Use Case

CISTO Experiences with an iRODS-Based Data Management Service

*NASA Goddard Space Flight Center
Computational & Information Sciences &
Technology Office (CISTO)*

Glenn Tamkin

CISTO Mission



- *Traditional*

- Enable scientists to increase their understanding of the Earth and the universe by providing state-of-the-art high performance computing, storage, network, and application solutions
- Provide large-scale compute engines, analytics, data sharing, and high-end computing services

- *Future*

- Develop a data services capability to better support the climate research communities and prepare the way for technology advances



Challenges



- *Finding* observational and model data for use in climate and weather studies
- *Accessing* the geographically distributed data
- *Managing* the massive digital holdings, which are measured in petabytes and hundreds of millions of files
- *Maintaining* the data, which must often be preserved for decades
- *Supporting* data sharing, data publication, and data stewardship



Technology - integrated Rule-Oriented Data System (iRODS)



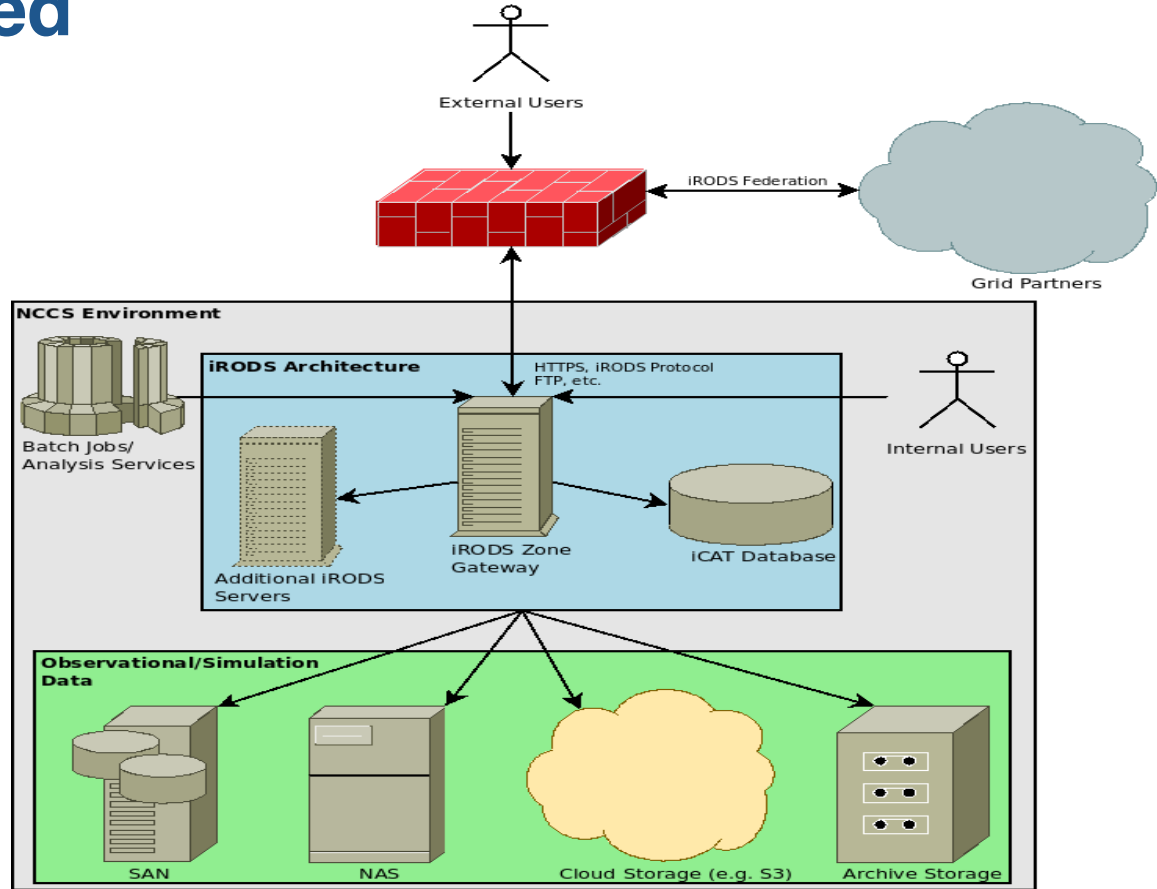
- Open source data grid software developed by the Data Intensive Cyber Environments (DICE) group, University of North Carolina
- Targets large repositories and digital preservation
- Supports the federation of independent, distributed collections
- Supports server-side workflows that are implemented by chaining execution rules together based on data policies
- Includes features such as domain-specific validation, automatic replication, and digital signature/checksum computation
- Validates assertions about data such as integrity and authenticity



CISTO Test Bed



- iRODS abstracts physical location of data
- iRODS assists with archive management



Ingest/Registration



- iRODS rules and microservices allow data to be stored in configurable collections based on data policies
- Replication to backup storage resources also supported

A screenshot of a web browser displaying the iRODS interface. The browser's address bar shows the URL 'https://169.154.148.17/irods/browse.php#ruri=rods@localhost%3A1...'. The interface includes a navigation pane on the left showing a tree view of collections: 'merra_Zone' > 'home' > 'public' > 'merra' > '1979'. The main pane displays a table of objects. The table has columns for 'Name', 'Resource', 'Size', and 'Date Modified'. The objects listed are all named 'MERRA100.prod.assim.instM_3d...' and are located in the 'demoResc' resource. The sizes range from 125.85 MB to 128.23 MB, and the dates are all from July 14, 2010.

Name	Resource	Size	Date Modified
MERRA100.prod.assim.instM_3d...	demoResc	125.88 MB	July 14, 2010, 11:02 am
MERRA100.prod.assim.instM_3d...	demoResc	126.24 MB	July 14, 2010, 11:07 am
MERRA100.prod.assim.instM_3d...	demoResc	127.29 MB	July 14, 2010, 11:11 am
MERRA100.prod.assim.instM_3d...	demoResc	128.23 MB	July 14, 2010, 11:15 am
MERRA100.prod.assim.instM_3d...	demoResc	126.73 MB	July 14, 2010, 11:19 am
MERRA100.prod.assim.instM_3d...	demoResc	124.78 MB	July 14, 2010, 11:29 am
MERRA100.prod.assim.instM_3d...	demoResc	125.34 MB	July 14, 2010, 11:33 am
MERRA100.prod.assim.instM_3d...	demoResc	126.65 MB	July 14, 2010, 11:37 am
MERRA100.prod.assim.instM_3d...	demoResc	128.45 MB	July 14, 2010, 11:41 am
MERRA100.prod.assim.instM_3d...	demoResc	127.45 MB	July 14, 2010, 11:46 am
MERRA100.prod.assim.instM_3d...	demoResc	125.85 MB	July 14, 2010, 11:50 am

Search



- iRODS rules and microservices can be used to assign metadata
- iRODS provides advanced search capabilities over the metadata

The screenshot displays the iRODS web interface. An 'Advanced Search' dialog box is open, showing search criteria for a file. The search results table below the dialog shows the following metadata:

Attribute	Value
variables	like Surface Geopotential
checksum	= 67abb8ca2e184ad2f92
Name	Op Value
Name	Op Value
Name	Op Value

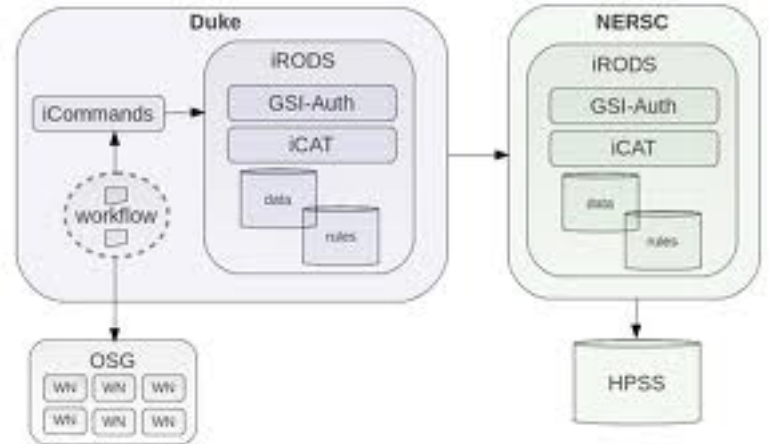
The metadata table also includes the following information:

- name: HDFEOS_V2.14
- dimensions: TIME:EOSGRID = 1, YDim:EOSGRID = 144, XDim:EOSGRID = 288, Height:EOSGRID = 42
- conventions: CF-1.0
- contact: http://gmao.gsfc.nasa.gov/
- comment: GEOS-5.2.0
- checksum: 91ddec7eee867abb8ca2e184ad2f92

Federation



- Federation allows separate iRODS Zones (iRODS instances) to become integrated.
- Each zone is a separate iRODS instance, administered separately, but the users in the multiple zones, if given permission, will be able to access data and metadata in the other zones.
- No user passwords are exchanged, as each system will, in a secure manner, check with the user's local zone for authentication when the user connects.



Preliminary Tests – Observational Data

- Developed an iRODS data grid that publishes Moderate Resolution Imaging Spectroradiometer (MODIS) observational data
 - 54 million registered files, 630 TB of data, and over 300 million defined metadata values
- Developed an iRODS data grid that focuses on a small-scale, multi-product, application-specific data service
 - The Invasive Species Data Service (ISDS) manages a collection of MODIS data products for ecological forecasting applications



modis_Zone



Collection	MODIS Atmosphere
Data	Aerosol, Water Vapor, Cloud, Profile, Cloud Mask, Joint Products
Type	Observational
Format	HDF
Customers	GES DISC, MODIS Science Data Support Team / Admins, Users
Distinction	Operational environment, 40,000,000 Data Objects!
Interfaces	Programmatic (Admin), FUSE (User), iRODS clients
Status	TRL 3 => TRL 6 (Subsystem validation in an operational environment.)

One of the most important ecological issues concerning our planet is climate change. It is generally agreed that the Earth's climate will modify in response to radiative forcing induced by changes in atmospheric trace gases, cloud cover, cloud type, solar radiation, and other factors.



isds_Zone



Collection	Invasive Species Data Service (ISDS)
Data	MODIS Land NDVI Phenology (Time-Series) Data
Type	Observational
Format	GeoTIFF
Customers	MD DNR, DOI BLM (GSENM) / Users
Distinction	Personal-/laboratory-scale, application-specific (ISFS) iRODS-based DMS; ISFS/ISDS licensed for distribution
Interfaces	isds_CI (User), iRODS clients
Status	TRL 3 => TRL 7 (System validation in an operational environment.)

The Invasive Species Forecasting System (ISFS) is a modeling framework that allows users to load point occurrence field sample data for a plant species of interest and quickly generate habitat suitability maps for geographic regions of management concern, such as a national park, monument, forest, or refuge. Target customers for applications built using ISFS are natural resource managers and decision-makers who have a need for scientifically valid, model-based predictions of the habitat suitability of plant species of management concern.


Invasive Species
Forecasting System




Preliminary Tests – Simulation Data



- Developed an iRODS data grid that manages Modern Era Retrospective-Analysis for Research and Applications (MERRA) simulation data
 - 360 files, 47 GB of data, and 4000 metadata values
- Developed an iRODS data grid that publishes Year of Tropical Convection (YOTC) data sets
 - 134,000 files, 12 TB of data, and 400,000 metadata values




merra_Zone




Collection	Modern Era Retrospective-Analysis for Research and Applications (MERRA)
Data	Monthly products from the past 15 years
Type	Observational/Simulation
Format	NETCDF
Customers	NCCS, GES DISC, ESG, Nebula / Admins, Managers
Distinction	merra_Zone @ NCCS + merra_Zone @ Nebula
Interfaces	merra_CI (Admin), iRODS clients
Status	TRL 3 => TRL 5 <i>(System validation in a relevant test environment.)</i>

Retrospective-analyses (or reanalyses) have been a critical tool in studying weather and climate variability for the last 15 years. Reanalyses combine observational data with model output to provide a consistent and comprehensive picture of the Earth's atmosphere, land, and oceans.




yotc_Zone



Collection	Year of Tropical Convection (YOTC)
Data	Satellite, in-situ and simulation/prediction model data sets
Type	Observational/Simulation
Format	NETCDF
Customers	NCCS / Admins, Users
Distinction	Operational environment, iRODS-mediated archive management
Interfaces	yotc_CI (Admin), FUSE (User), iRODS clients
Status	TRL 3 => TRL 7 <i>(System validation in an operational environment.)</i>

The realistic representation of tropical convection in our global atmospheric models is a long-standing grand challenge for numerical weather forecasts and global climate predictions. To address the challenge of tropical convection, collaborative organizations from around the world have proposed a year of coordinated observing, modeling and forecasting of organized tropical convection and its influences on predictability. This effort is intended to exploit the vast amounts of existing and emerging observations, the expanding computational resources and the development of new, high-resolution modeling frameworks, with the objective of advancing the characterization, diagnosis, modeling, parameterization and prediction of multi-scale convective/dynamic interactions, including the two-way interaction between tropical and extra-tropical weather/climate.



NASA LARC/GSFC iRODS Federation



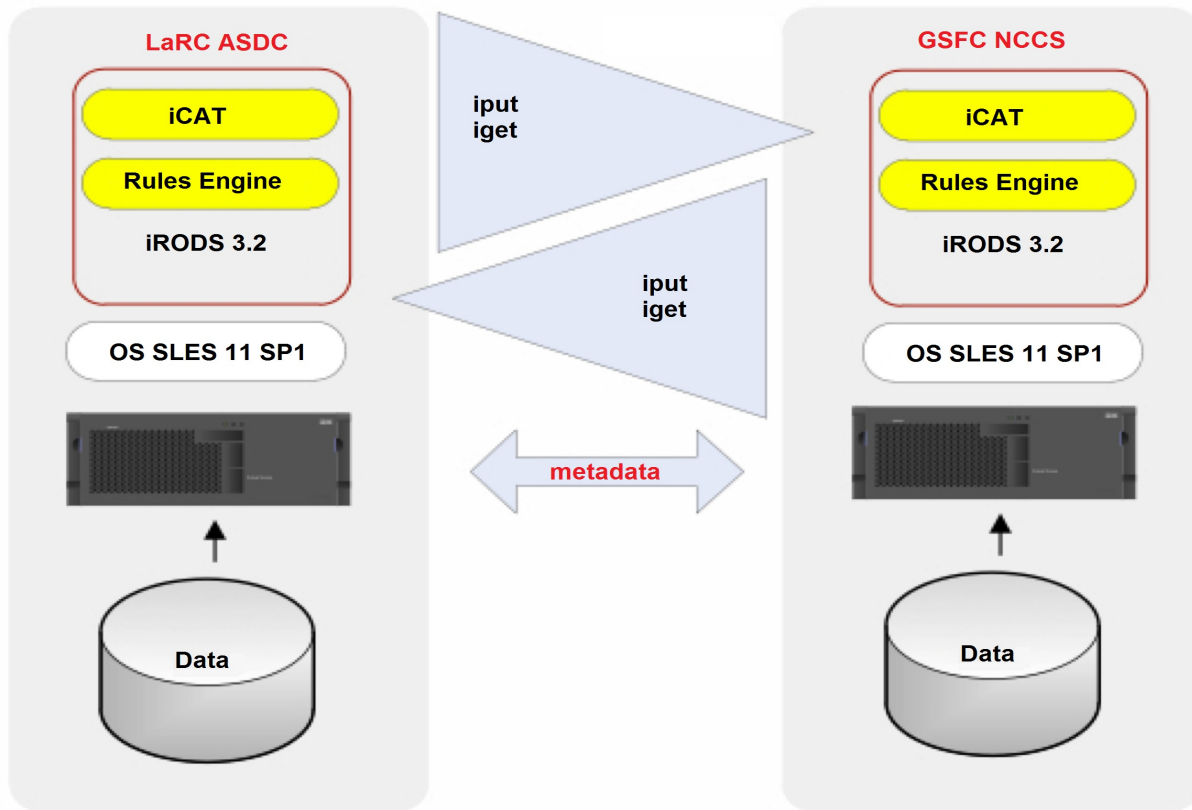
- Stakeholders

- NASA's Atmospheric Science Data Center (ASDC) and Center for Climate Simulation (NCCS) both ingest, archive, and distribute data that is essential to stakeholders including the climate research community, science applications community, and a growing community of government and private-sector customers who have a need for atmospheric and climatic data.

- Goals

- To implement a data federation ability to improve and automate the discovery of heterogeneous data, decrease data transfer latency, and meet customizable criteria based on data content, data quality, metadata, and production. Another goal of the federation is to support/enable applications and customers that require the integration of multiple heterogeneous data collections.

NASA LARC/GSFC Federation Architecture



Langley Research Center (LARC) Zone



- The LARC Atmospheric Science Data Center (ASDC) Zone contains various scientific collections (e.g., CALIPSO, CERES)
- ASDC's Data Products Online (DPO) is federated with the GSFC/MAS Zone.

Name	Resource	Size	Date Modified
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-01T01-09-20ZN.hdf.met	DPOResc	5.4 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-01T06-06-01ZN.hdf.met	DPOResc	5.4 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-01T09-23-46ZN.hdf	DPOResc	39.68 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-01T18-24-42ZD.hdf	DPOResc	44.7 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-02T05-10-34ZN.hdf	DPOResc	39.49 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-02T07-35-49ZD.hdf	DPOResc	44.7 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-02T13-25-00ZN.hdf.met	DPOResc	5.38 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-02T15-03-55ZN.hdf.met	DPOResc	5.4 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-02T15-50-20ZD.hdf.met	DPOResc	5.41 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-02T17-29-15ZD.hdf.met	DPOResc	5.41 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-02T18-21-41ZN.hdf	DPOResc	39.68 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-02T20-00-36ZN.hdf	DPOResc	39.68 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T02-36-07ZN.hdf	DPOResc	39.68 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T02-36-07ZN.hdf.met	DPOResc	5.41 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T05-01-27ZD.hdf	DPOResc	44.7 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T06-40-22ZD.hdf.met	DPOResc	5.38 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T07-32-47ZN.hdf	DPOResc	39.68 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T09-58-08ZD.hdf	DPOResc	44.7 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T10-50-38ZN.hdf.met	DPOResc	5.42 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T14-54-48ZD.hdf	DPOResc	44.7 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T15-47-18ZN.hdf	DPOResc	39.49 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T15-47-18ZN.hdf.met	DPOResc	5.39 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T21-30-24ZD.hdf.met	DPOResc	5.42 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-03T22-22-49ZN.hdf.met	DPOResc	5.4 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-04T03-19-30ZN.hdf.met	DPOResc	5.42 KB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-04T04-06-00ZD.hdf	DPOResc	44.7 MB	March 19, 2014, 10:57
CAL_LID_L2_VFM-ValStage1-V3-30.2013-04-04T04-06-00ZD.hdf.met	DPOResc	5.4 KB	March 19, 2014, 10:57

Goddard Space Flight Center (GSFC) Zone



- The GSFC MERRA Analytic Service (MAS) Zone contains the full MERRA collection
- All NetCDF metadata is registered for each file
- Modern-Era Retrospective Analysis for Research and Application (MERRA) is federated with the LARC/ASDC Zone

The screenshot shows a file browser interface. On the left, a directory tree is visible under 'masZone', showing folders for 'home', 'anon', and 'MERRA'. The 'MERRA' folder contains subfolders for various MERRA versions (e.g., MA1NXINT.5.2.0, MAI3CPASM.5.2.0, etc.) and a folder for the year '1979'. On the right, a table displays a list of files. The table has columns for 'Name', 'Resource', 'Size', and 'Date Modified'. The files listed are MERRA100.prod.assim.tavg3_3d_chm_Nv.19790131.hdf through MERRA100.prod.assim.tavg3_3d_chm_Nv.19790108.hdf. The file MERRA100.prod.assim.tavg3_3d_chm_Nv.19790114.hdf is highlighted in blue. The status bar at the bottom indicates 'Page 1 of 1' and 'Displaying objects 1 - 31 of 31'.

Name	Resource	Size	Date Modified
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790131.hdf	attDAR01	881.93 MB	August 29, 2013, 3:01 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790130.hdf	attDAR01	886.07 MB	August 29, 2013, 3:01 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790128.hdf	attDAR01	890.13 MB	August 29, 2013, 3:01 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790129.hdf	attDAR01	891.31 MB	August 29, 2013, 3:01 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790127.hdf	attDAR01	895.69 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790125.hdf	attDAR01	890.08 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790126.hdf	attDAR01	895.87 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790124.hdf	attDAR01	883.75 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790122.hdf	attDAR01	880.17 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790123.hdf	attDAR01	883.16 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790121.hdf	attDAR01	877.53 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790120.hdf	attDAR01	876.16 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790119.hdf	attDAR01	882.27 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790118.hdf	attDAR01	882.86 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790116.hdf	attDAR01	878.99 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790117.hdf	attDAR01	879.83 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790115.hdf	attDAR01	872.23 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790113.hdf	attDAR01	875.18 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790114.hdf	attDAR01	875.81 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790112.hdf	attDAR01	874.69 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790111.hdf	attDAR01	876.38 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790110.hdf	attDAR01	873.54 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790109.hdf	attDAR01	869.13 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790108.hdf	attDAR01	875.74 MB	August 29, 2013, 2:58 am

LARC/GSFC Federation



- Both iRODS Zones are accessible together
- Users can access the data via web browser, command line, FUSE file system, etc.

The screenshot displays the iRODS web interface. On the left, a 'Collections' tree shows a hierarchy starting with 'asdcZone', followed by 'home', 'public', 'rods', and 'masZone'. The 'masZone' folder is expanded to show sub-folders like 'home' and 'anon'. The main area shows a table of files with columns for Name, Resource, Size, and Date Modified. The table lists numerous files, all with the resource 'attDAR01'. The file 'MERRA100.prod.assim.tavg3_3d_chm_Nv.19790114.hdf' is highlighted in blue. At the bottom, the page indicates 'Page 1 of 1' and 'Displaying objects 1 - 31 of 31'.

Name	Resource	Size	Date Modified
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790131.hdf	attDAR01	881.93 MB	August 29, 2013, 3:01 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790130.hdf	attDAR01	886.07 MB	August 29, 2013, 3:01 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790128.hdf	attDAR01	890.13 MB	August 29, 2013, 3:01 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790129.hdf	attDAR01	891.31 MB	August 29, 2013, 3:01 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790127.hdf	attDAR01	895.69 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790125.hdf	attDAR01	890.08 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790126.hdf	attDAR01	895.87 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790124.hdf	attDAR01	883.75 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790122.hdf	attDAR01	880.17 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790123.hdf	attDAR01	883.16 MB	August 29, 2013, 3:00 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790121.hdf	attDAR01	877.53 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790120.hdf	attDAR01	876.16 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790119.hdf	attDAR01	882.27 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790118.hdf	attDAR01	882.86 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790116.hdf	attDAR01	878.99 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790117.hdf	attDAR01	879.83 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790115.hdf	attDAR01	872.23 MB	August 29, 2013, 2:59 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790113.hdf	attDAR01	875.18 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790114.hdf	attDAR01	875.81 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790112.hdf	attDAR01	874.69 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790111.hdf	attDAR01	876.38 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790110.hdf	attDAR01	873.54 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790109.hdf	attDAR01	869.13 MB	August 29, 2013, 2:58 am
MERRA100.prod.assim.tavg3_3d_chm_Nv.19790108.hdf	attDAR01	875.74 MB	August 29, 2013, 2:58 am

LARC/GSFC Metadata



- All NetCDF files are harvested for metadata at registration time using iRODS microservices and rules
- Users can view and search across the metadata
- Custom/derived metadata can be applied to any collection or file

A screenshot of the iRODS metadata viewer interface. The top window shows a file list with columns for Name, Resource, Size, and Date Modified. The selected file is 'MERRA100.prod.assim.tavg3_3d_chm_Nv.19790131.hdf'. A secondary window displays the metadata for this file in a table with columns for Name, Value, and Unit. The metadata includes various global and dataset-specific attributes such as title, source, references, and dimensions.

Name	Value	Unit
var:DTRAIN:missing_value	1.e+15f	
var:DTRAIN:long_name	Detrainment cloud mass flux	
var:DTRAIN:fmissing_value	1.e+15f	
var:DTRAIN:dim	TIME\EOSGRID, Height\EOSGRID, YDim\EOSGRID, XDim\EOSGRID	
var:DTRAIN:add_offset	0.f	
glb:title	MERRA reanalysis. GEOS-5.2.0	
glb:StructMetadata.0	GROUP	
glb:source	Global Modeling and Assimilation Office. GEOSops_5_2_0	
glb:references	http://gmao.gsfc.nasa.gov/research/merra/	
glb:missing_value	1.e+15f	
glb:list	HDFEOSVersion, StructMetadata.0, missing_value, Conventions, title, history, in	
glb:institution	Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Gr	
glb:history	File written by CFIO	
glb:HDFEOSVersion	HDFEOS_V2.14	
glb:CoreMetadata.0	GROUP	
glb:Conventions	CF-1.0	
glb:contact	http://gmao.gsfc.nasa.gov/	
glb:comment	GEOS-5.2.0	
glb:ArchivedMetadata.0	GROUP	
dim:YDim\EOSGRID	361	
dim:XDim\EOSGRID	540	
dim:TIME\EOSGRID	8	
dim:list	TIME\EOSGRID, Height\EOSGRID, YDim\EOSGRID, XDim\EOSGRID	
dim:Height\EOSGRID	72	
checksum	85d1171d54538e7784280ce9d150a87f	

LARC/GSFC Search



- Files can be searched by attribute or metadata

Advanced Search

Attributes

Name:

Modified Within:

Owner:

Resource:

Only: Under Current Collection

Current Collection:

Metadata

<input type="text" value="var:VC:long_name"/>	<input "="" type="text" value="="/>	<input type="text" value="Northward component of v"/>
<input type="text" value="Name"/>	<input type="text" value="Op"/>	<input type="text" value="Value"/>
<input type="text" value="Name"/>	<input type="text" value="Op"/>	<input type="text" value="Value"/>
<input type="text" value="Name"/>	<input type="text" value="Op"/>	<input type="text" value="Value"/>
<input type="text" value="Name"/>	<input type="text" value="Op"/>	<input type="text" value="Value"/>

MA13CPASM.5.2.0

MA16NVANA.5.2.0

Summary

- iRODS federation can be a powerful mechanism for creating ensembles of distributed data
- Collections are still managed by the data owners, but iRODS provides mechanisms for creating and enforcing policies that control external access within the federation
- iRODS abstracts the advertised collection and the physical location of the data. As a result, the data can be relocated or extended without disturbing the federated consumer. In fact the consumer needn't be aware of the federation, as it appears as one large ensemble
- This approach enables parallel downloads of datasets from selected replica servers that can be geographically dispersed, but still accessible by users worldwide

Next Steps

- Inclusion of a new member to the federation
- Extended access to files on the ASDC's tape archive
- Collection expansion to include LandSAT and MODIS data
- Utilization of the NetCDF plugin to assign metadata to arbitrary NetCDF/HDF4/HDF5 data.
- Testing/utilization of the iRODS HDFS plugin (Hadoop access)
- Testing/utilization of methods to register remote data in the Public Domain via iRODS MicroService Objects
- Automated file registration using operating system level triggers (i.e., 'iNotify')

Future ASDC iRODS Configuration

