iRODS and Large-Scale Data Management

Leesa Brieger

renci

RESEARCH \ ENGAGEMENT \ INNOVATION

Renaissance Computing Institute
(RENCI)

- A research unit of UNC Chapel Hill
- Current Director: Stan Ahalt, formerly from the Ohio Supercomputer Center
- State-supported
- Governed by the Triangle universities:
  - UNC Chapel Hill
  - NC State University
  - Duke University

Data Intensive Cyber Environments

(DICE Group)

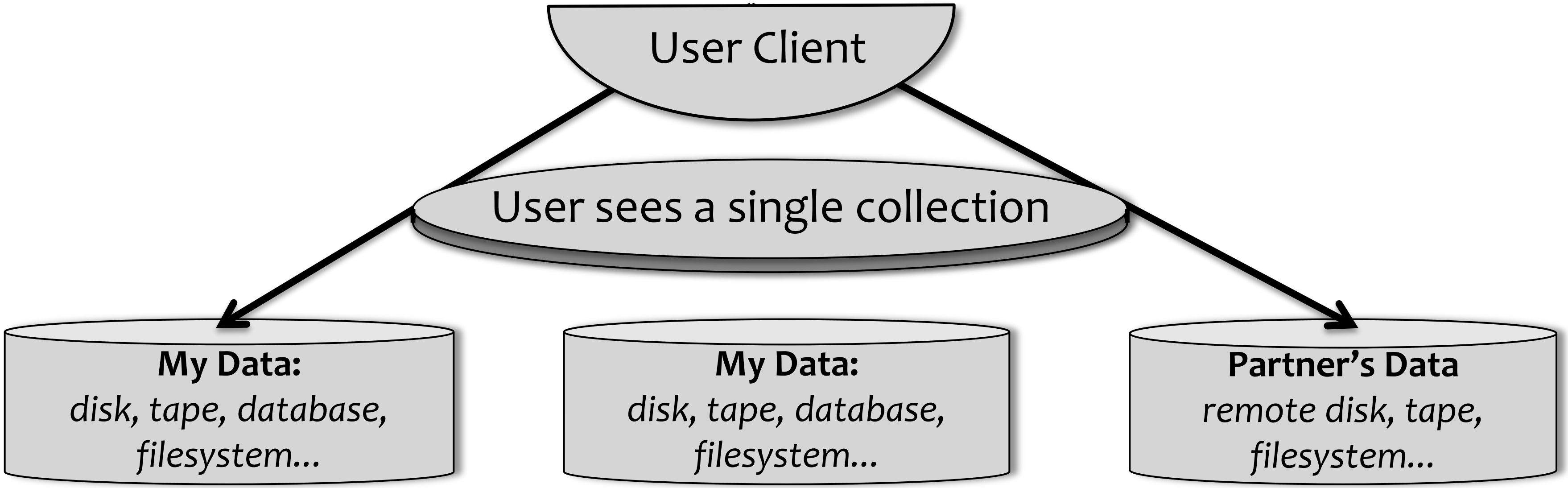
- Directed by Reagan Moore
- Developed SRB, the Storage Resource Broker
- Began at SDSC, the San Diego Supercomputer Center
- Most of the group migrated to UNC Chapel Hill in 2008
  - Joint appointments in SILS and RENCI
  - The group is bi-coastal: DICE-UNC, DICE-UCSD
- SRB migrated to iRODS, the integrated Rule-Oriented Data System in 2009

iRODS Evolution

- Based on decade-long SRB development experience for managing distributed data
- Community-driven
- iRODS picked up where SRB left off
- Modular, extensible, customizable
- Open source (BSD license – Berkeley Software Distribution)
- Supported by RENCI at UNC: iRODS@RENCI

iRODS Unified Virtual Collection

iRODS View of Distributed Data


User Client
User sees a single collection
My Data:
disk, tape, database, filesystem...
My Data:
disk, tape, database, filesystem...
Partner's Data
remote disk, tape, filesystem...

iRODS installs over heterogeneous data resources

Users share & manage distributed data as a single collection

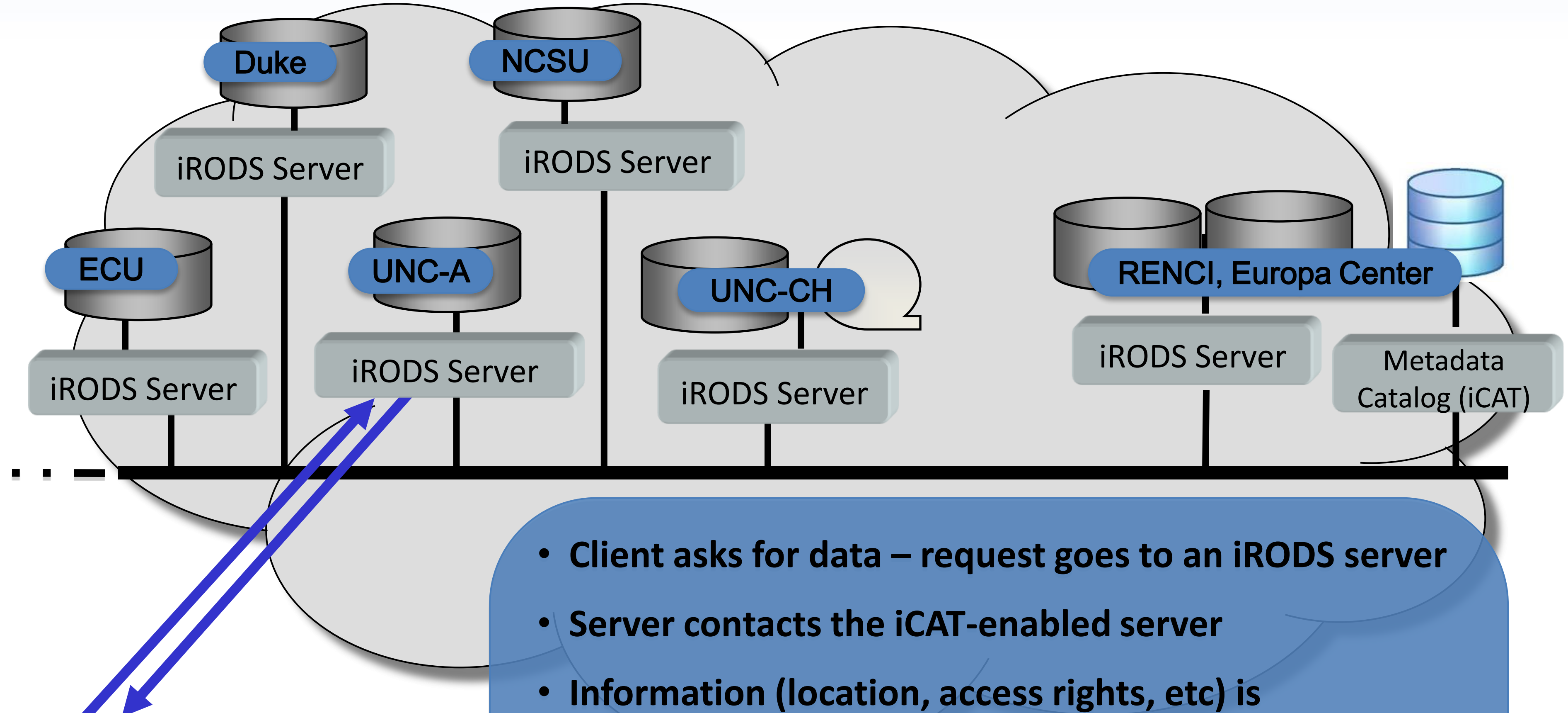
renci

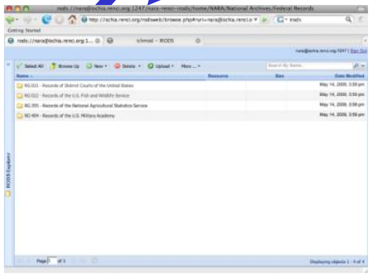
iRODS as a Data Grid

- Sharing data across:
  - geographic and institutional boundaries
  - heterogeneous resources (hardware/software)
- Virtual collections of distributed data
- Global name spaces
  - data/files
  - users: single sign on
  - storage: virtual resources
- Metadata catalogue (iCAT) manages mappings between logical and physical name spaces
- Beyond a single-site repository model

renci

A RENCI Data Grid



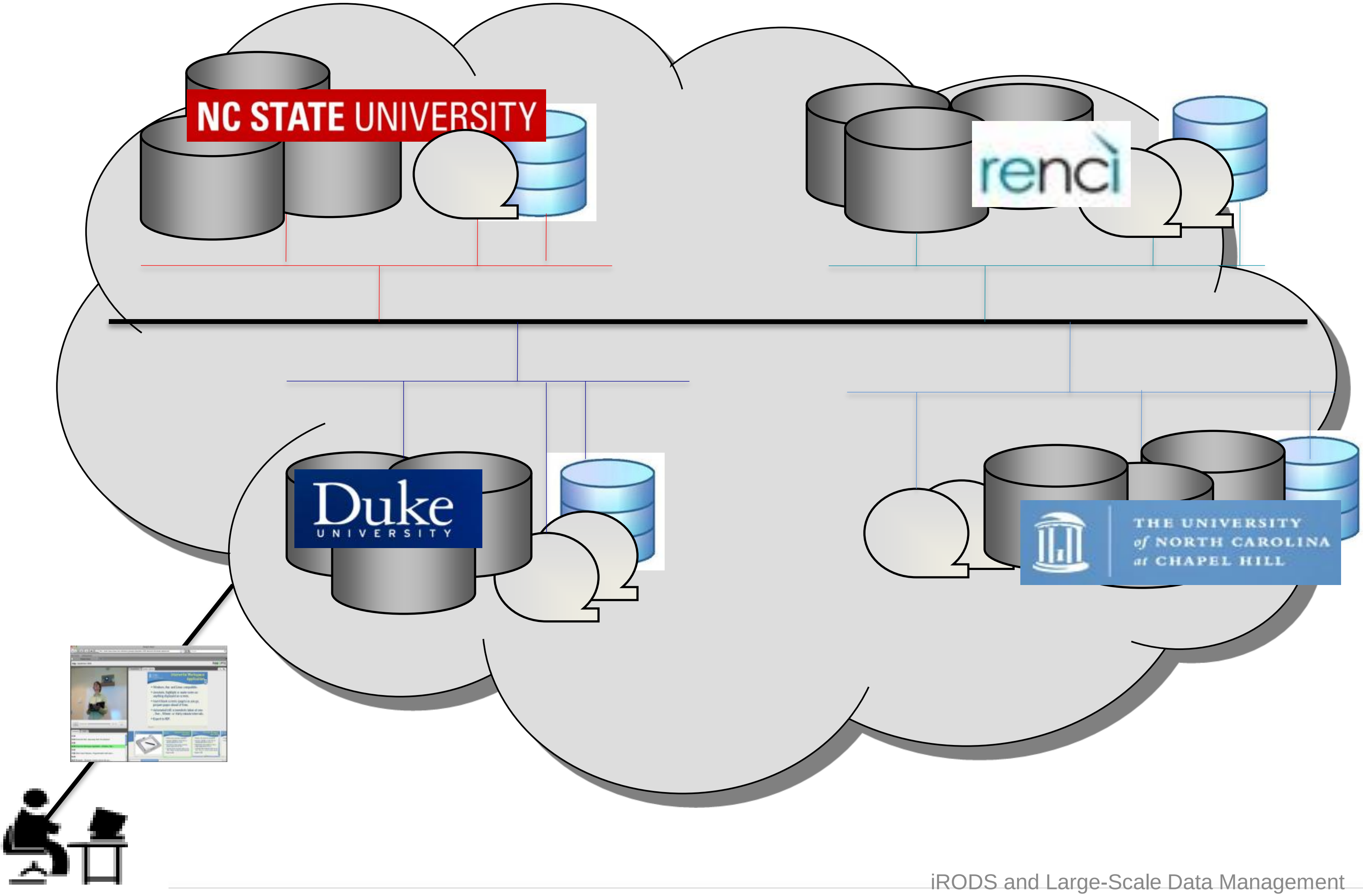

rods://nara@ischia.renci.org:1247/nara-renci-irods/home/NARA/National Archives/Federal Records
http://ischia.renci.org/rodsweb/browse.php#ruri=nara@ischia.renci.o
irods
Getting Started
rods://nara@ischia.renci.org:1...
ichmod - iRODS
nara@ischia.renci.org:1247 | Sign Out
Select All
Browse Up
New
Delete
Upload
More ...
Search By Name...
iRODS Explorer

Name	Resource	Size	Date Modified
RG 021 - Records of District Courts of the United States			May 14, 2008, 3:59 pm
RG 022 - Records of the U.S. Fish and Wildlife Service			May 14, 2008, 3:59 pm
RG 355 - Records of the National Agricultural Statistics Service			May 14, 2008, 3:59 pm
RG 404 - Records of the U.S. Military Academy			May 14, 2008, 3:59 pm

- Client asks for data – request goes to an iRODS server
- Server contacts the iCAT-enabled server
- Information (location, access rights, etc) is retrieved from the iCAT
- Server containing data is signaled to send data to authorized client

TUCASI Infrastructure Project (TIP)

Federated Data Grids



UNC iRODS Data Grids, Federations, and Collaborations

THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

NARA CIBER------------------

NCDC (NOAA)---------------

NCCS (NASA)----------------

MotifNet----------------------

CineGrid----------------------

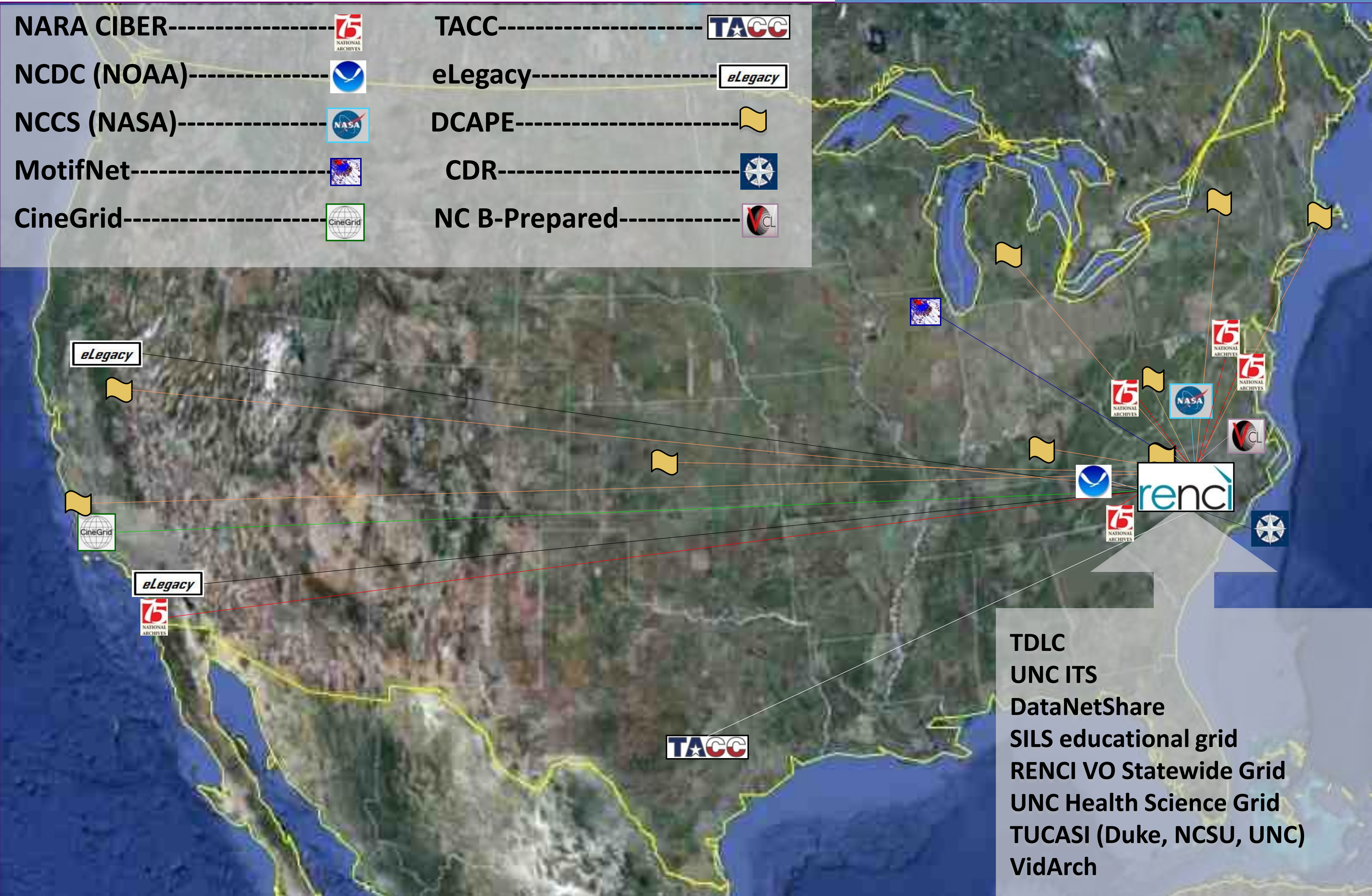
TACC-----------------------

eLegacy--------------------

DCAPE-------------------------

CDR---------------------------

NC B-Prepared-------------



TDLC
UNC ITS
DataNetShare
SILS educational grid
RENCI VO Statewide Grid
UNC Health Science Grid
TUCASI (Duke, NCSU, UNC)
VidArch

The Data Issues - Examples

Genome centers

- petabytes of data
- researchers sharing data
- derived data products from workflows
- requirements for traceability and reproducibility (provenance and metadata management)

NOAA’s National Climate Data Center (NCDC)

- repository management
- publishing public data
- delivery of services with the data (to the public and to researchers)
- support for climate modeling (at ORNL, …)

How much is that?

1 kilobyte (kB) = 1000 bytes
1 megabyte (MB) = 1000 KB
1 gigabyte (GB) = 1000 MB
1 terabyte (TB) = 1000 GB
1 petabyte (PB) = 1000 TB

- 1 byte = 8 bits
- 1 kB = a joke (very short story)
- 2 kB = 1 typewritten page
- 1 MB = 4 books
- 2 MB = 1 high-resolution photo
- 5 MB = complete works of Shakespeare
- 500 MB = a CD-ROM
- 1 GB = ~ 900,000 pages of plaintext (enough typewritten pages to fill the bed of a pickup)
- 2 GB = 20 yards of books on a shelf
- 1 TB = 900+ million pages of plaintext(4.5 million books); ~230 DVD movies
- 20 TB = entire printed collection of Library of Congress
- 1 PB = 4.7 billion books; 13.3 years of HD-TV video
- 1.5PB = 10 billion photos on Facebook
- 50 PB = entire written works of mankind from the beginning of recorded history, in all languages

The Issues – More Examples

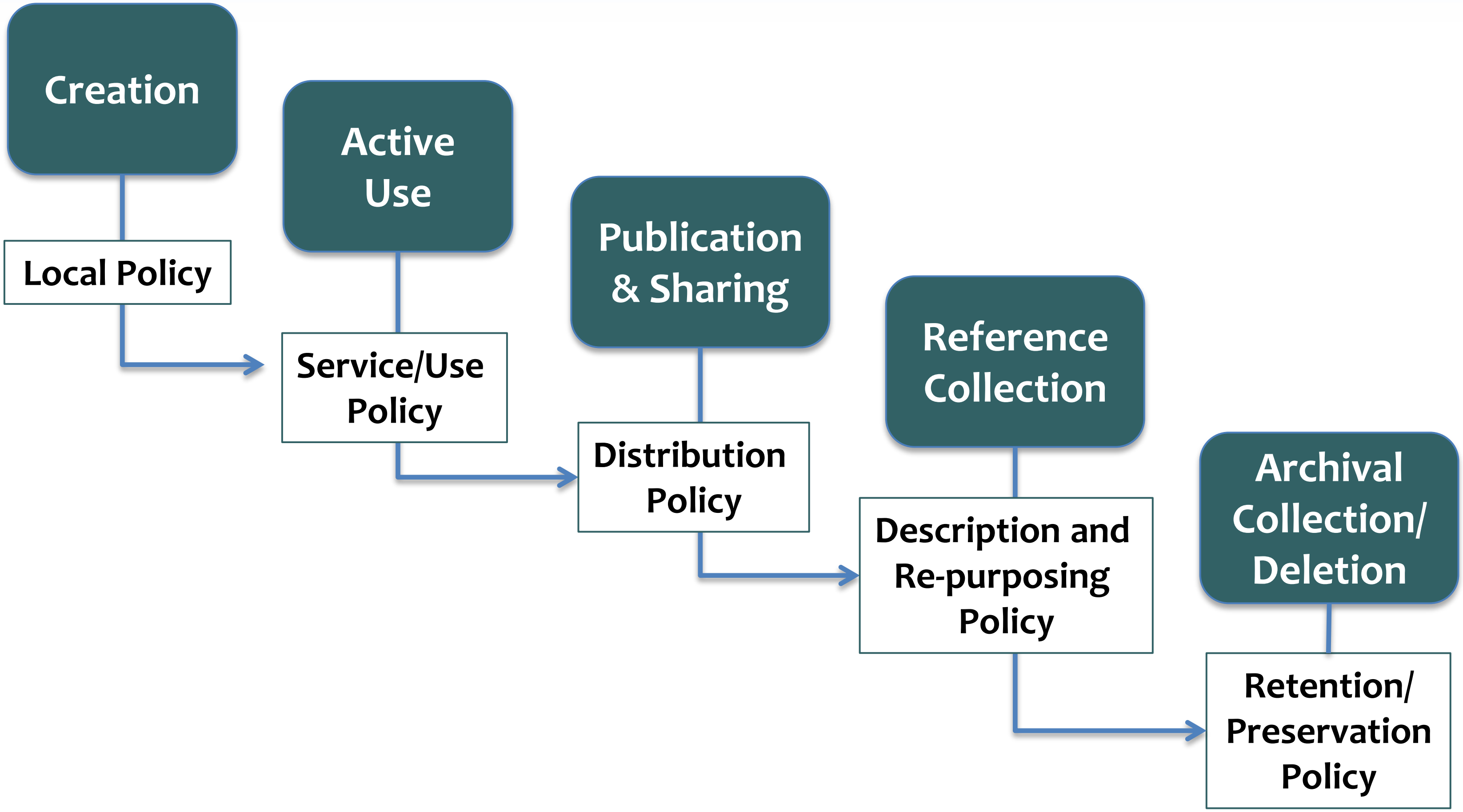
Streamline/automate data movement (HPC, data analysis)

- data to distributed jobs
- services provided on distributed data
- collect data to a central location for sharing/post-processing
- archive data output

National/agency repositories

- publish data from a multitude of disparate sources
- manage collections independently:
  - some public collections
  - some shared collections
  - varying life spans
  - some privacy-protected data (legal issues)
  - different requirements of integrity, versioning, provenance, etc [legal issues]

Data Life Cycle



Policy evolution across the stages of the data life cycle.

Policy-Based Data Environments

- Assembled/Distributed Collections
- Properties - attributes that ensure the *purpose* of the collections
- Policies - methodologies for enforcing desired *properties*
- Procedures - functions that implement the *policies*
  implemented as computer actionable rules/workflows
- Persistent state information - results of applying the *procedures*
  contained in system metadata
- Assessment criteria - validation that *state information* conforms to the desired *purpose*
  mapped to periodically executed policies

renci

Policy-Oriented Data Infrastructure

- Implement management policies
  - Each community defines their own policies

- Automate administrative tasks
  - Data administrator (not system administrator) manages data collections

- Validate assessment criteria
  - Verify policy compliance
    - Point-in-time through queries on metadata catalog
    - Compliance over time through parsing audit trails

Additional iRODS Design Goals

Virtualize distributed collections (data grid)

- Collection management independent of remote storage locations
- Infrastructure independence

Abstract out the data management

- Policy-based data management
- Separate policy enforcement from storage administration

Scalability and extensibility

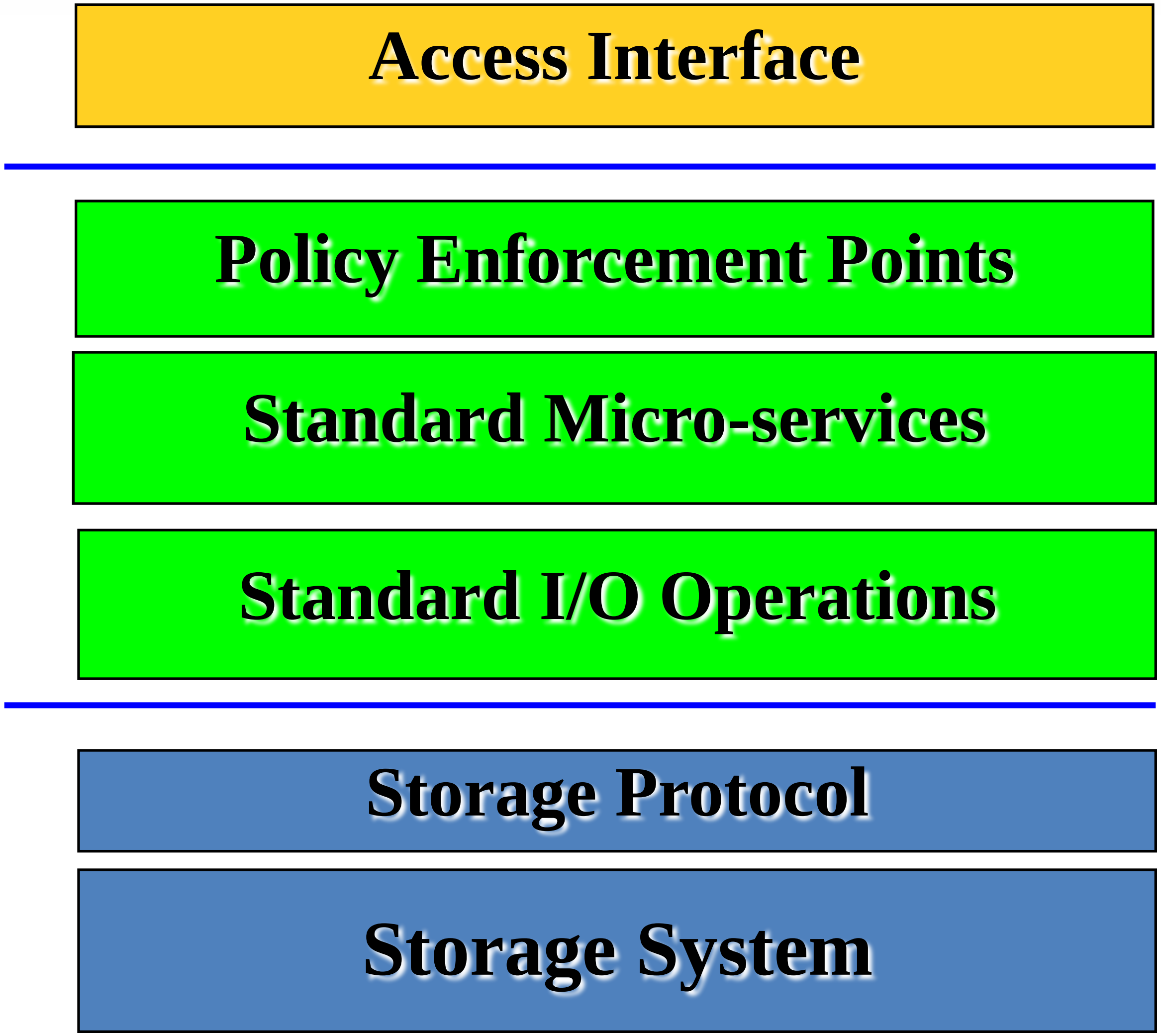
- Enable versioning of policies and procedures and data
- Support differentiated services (storage, metadata, messaging, workflow, scheduling)

iRODS Policy Implementation

Microservices and Rules

- Microservice – functional unit of work (C programs)
- Rules – workflows of microservices (and rules)
- Provide server-side (data-side) services
- Modular
- New libraries of microservices can be developed without touching the core code
- Allows customization of data grid for community-specific policy
- Event-triggered rule execution

Data Virtualization



Access Interface

Policy Enforcement Points

Standard Micro-services

Standard I/O Operations

Storage Protocol

Storage System

Map from the actions requested by the client to multiple policy enforcement points.

Map from policy to standard micro-services.

Map from micro-services to standard Posix I/O operations.

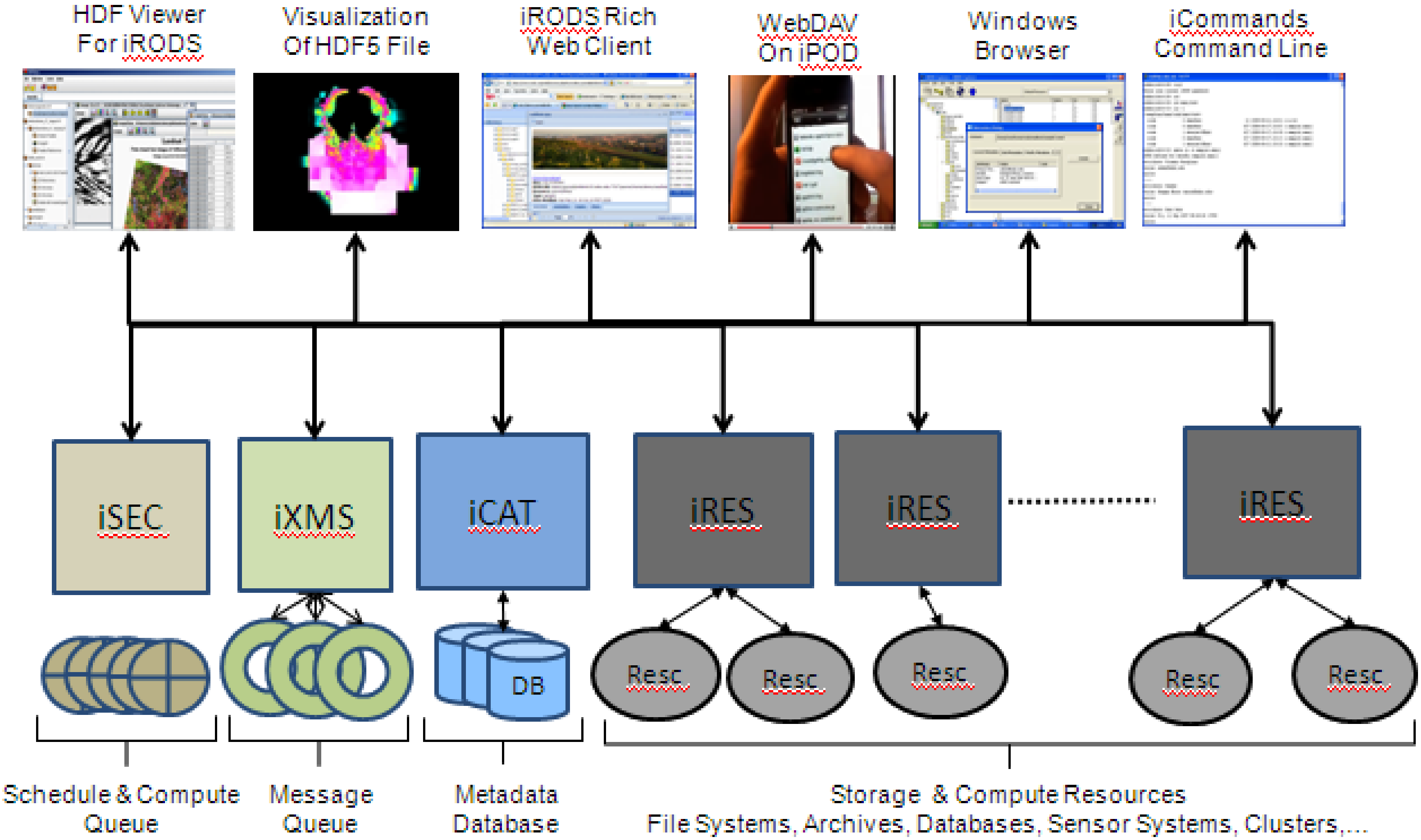
Map standard I/O operations to the protocol supported by the storage system

iRODS Extensible Infrastructure

- Some extensions handled by user groups
  - Clients
  - Policies
  - Procedures

- iRODS (extensible) core infrastructure:
  - Network transport
  - Authentication/Authorization
  - Distributed storage access
  - Remote/Data-side execution of services
  - Metadata management
  - Messaging
  - Rule engine

An iRODS Overview



Large-Scale Data with iRODS

- NASA
- Broad Institute
- Sanger Institute

NASA Center for Climate Simulation

NCCS: simulation tools and supercomputing resources for NASA missions and future scientific discoveries…

http://www.nccs.nasa.gov/

Part of NASA's High End Computing Program

Data Management System (DMS) Project: applied R&D aimed at moving iRODS data grid technology into practical use in the NCCS. It's about:

Technology Readiness, Operational Transfer, and Cultural Change ...

NCCS Mandates

- Major clients:
  - NASA's Global Modeling and Assimilation Office (GMAO)
  - Goddard Institute for Space Studies (GISS)
- Data management services and analytical tools in support of the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5), due between 2013 and 2014
- The tie between NASA modeling efforts and satellite missions: working to integrate model outputs and observational data
- NCCS keeping pace with increasing heterogeneity and distributed nature of Earth Science data

Climate Data Services

The story of climate data

•It's old, global in scale, and all of it is relevant.

•It's one of the largest and fastest growing classes of scientific data.

•The types and sources of climate data are becoming increasingly diverse.

•Climate research is becoming increasingly data centric.

Data services as a core NCCS mission

•Climate data services represent an effort to respond to the story of climate data.

•The concept is being defined.

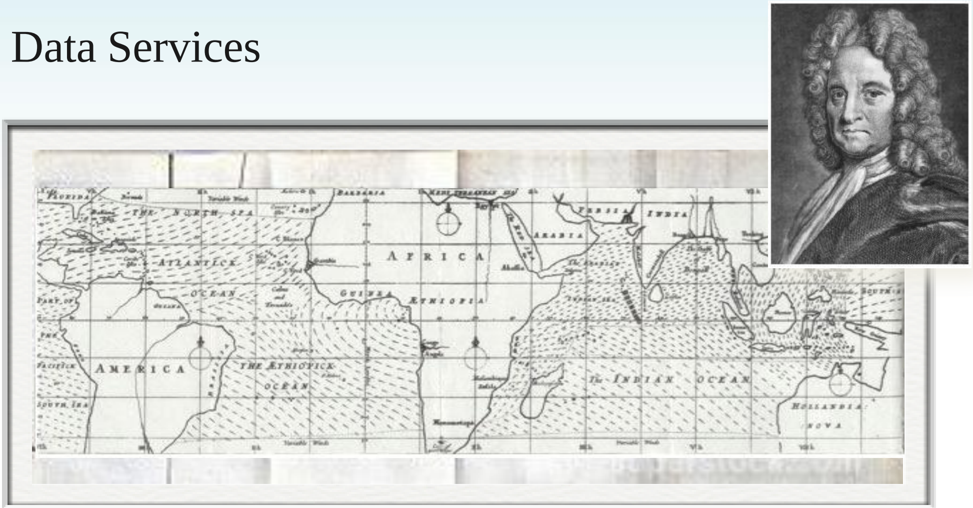
•At the very least, we know that it will require:

•a new view of data stewardship,

•new organizational processes, and

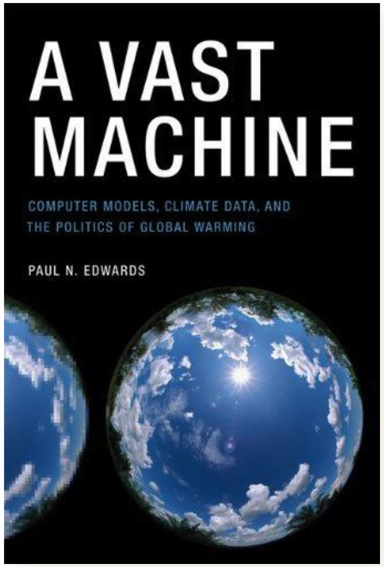
•new data technologies.

The DMS Project has been focused here ...



Edmund Halley's Wind Map (1686)

H.M.S. Challenger under Sail, 1874.



A VAST MACHINE

COMPUTER MODELS, CLIMATE DATA, AND THE POLITICS OF GLOBAL WARMING

PAUL N. EDWARDS

iRODS Capabilities

- ***Global persistent identifiers for naming digital objects***. A unique identifier is used for each object stored in iRODS. Replicas and versions of the same object share the same global identifier but differ in replication and version metadata. ***Support for metadata to identify system-level physical properties of the stored data object***. The properties that are stored include physical resource location, path names (or canned SQL queries in the case of database resources), owner, creation time, modification time, expiration times, file types, access controls, file size, location of replicas, aggregation in a container, etc. ***Support for descriptive metadata to enable discovery through simple query mechanisms***. iRODS supports metadata in terms of attribute-value-unit triplets. Any number of such associations can be added for each digital object. ***Standard access mechanisms***. Interfaces include Web browsers, Unix shell commands, Python load libraries, Java, C library calls, FUSE-based file interface, WebDav, Kepler and Taverna workflow, etc. ***Storage repository abstraction***. Files may be stored in multiple types of storage systems including tape systems, disk systems, databases, and, now, cloud storage. ***Inter-realm authentication***. The authentication system provides secure access to remote storage systems including secure passwords and certificate-based authentication such as Grid Security Infrastructure (GSI). ***Support for replication and synchronization of files between resource sites***. ***Support for caching copies of files onto a local storage system and support for accessing files in an archive using compound resource methodology***. This includes the concept of multiple replicas of an object with distinct usage models. Archives are used as "safe" copies and caches are used for immediate access. ***Support for physically aggregating files into tar-files to optimize management of large numbers of small files***.

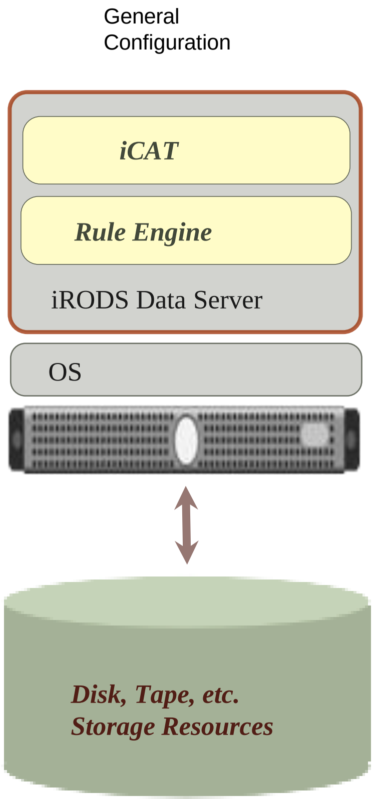
- ***Access controls and audit trails to control and track data usage***. ***Support for execution of remote operations for data sub-setting, metadata extraction, indexing, remote data movement, etc., using micro-services and rules***. ***Support for rich I/O models for file ingestion and access including* in situ *registration of files into the system, inline transfer of small files, and parallel transfer for large files***. ***Support for federation of data grids***. Two independently managed persistent archives can establish a defined level of trust for the exchange of materials and access in one or both directions. This concept is very useful for developing a full-fledged preservation environment with dark and light archives.

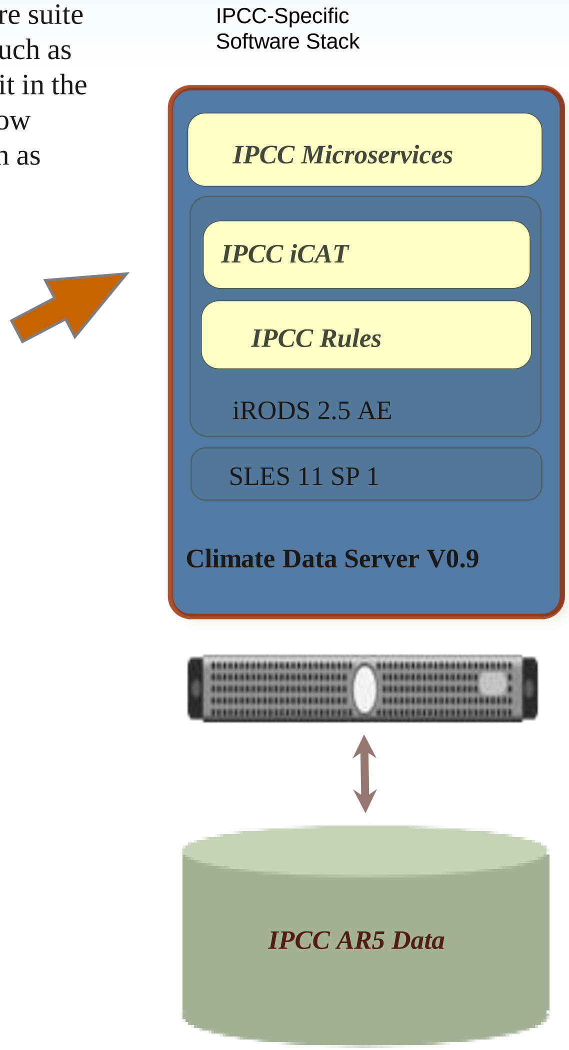
NASA iRODS Zones

- modis_Zone: MODIS Earth Observational Data
  - use iRODS to provide access to Moderate Resolution Imaging Spectroradiometer (MODIS) atmosphere data products (satellite data)
  - iRODS in the setting of a production, flight mission observational data system
  - 22 iRODS data servers, 54 million files, 300 million metadata attributes, 630 TB
- merra_Zone: MERRA Climate Simulation Data
  - use iRODS to provide access to Modern Era Retrospective-Analysis for Research and Applications (MERRA) model output data.
  - managing GMAO model output (simulation) data products.
- yotc_Zone: YOTC Climate Simulation Data
  - Production access to Year of Tropical Convection (YOTC) data (model output data)
- isds_Zone: Invasive Species Data Service (ISDS) Earth Observational Data
  - deliver custom data products
  - personal- or laboratory-scale data management, which could participate in an extended federation of iRODS resources

Climate Data Server (CDS) Anatomy

Our approach has been to build a core suite of general purpose scientific kits – such as NetCDF, HDF, and GeoTIF – that sit in the vertical stack above iRODS and below application-specific climate kits such as IPCC, MERRA, and SMOS ...





CDS Core Components

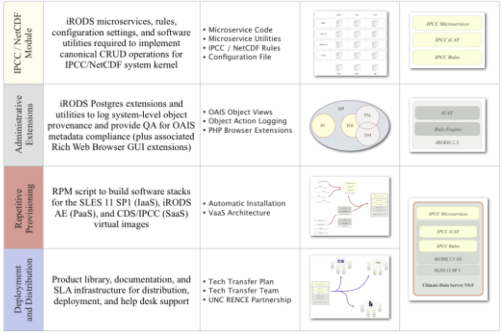
•IPCC-specific **microservices**
Canonical archive ops, particularly the mechanisms required to ingest OAIS-compliant Submission Information Package (SIP) metadata for IPCC NetCDF objects ...

•IPCC-specific **metadata**
OAIS-compliant constitutive (application-independent) Representation Information (RI) and Preservation Description Information (PDI) metadata for IPCC NetCDF objects ...

•IPCC-specific **rules**
IPCC NetCDF triggers and workflows ...

•A specific release of **iRODS**
iRODS 2.5 w/ Administrative Extensions (AE)

•A specific **operating system**
SLES 11 SP 1



IPCC / NetCDF Module	iRODS microservices, rules, configuration settings, and software utilities required to implement canonical CRUD operations for IPCC/NetCDF system kernel	• Microservice Code • Microservice Utilities • IPCC / NetCDF Rules • Configuration File
Administrative Extensions	iRODS Postgres extensions and utilities to log system-level object provenance and provide QA for OAIS metadata compliance (plus associated Rich Web Browser GUI extensions)	• OAIS Object Views • Object Action Logging • PHP Browser Extensions
Repetitive Provisioning	RPM script to build software stacks for the SLES 11 SP1 (IaaS), iRODS AE (PaaS), and CDS/IPCC (SaaS) virtual images	• Automatic Installation • VaaS Architecture
Deployment and Distribution	Product library, documentation, and SLA infrastructure for distribution, deployment, and help desk support	• Tech Transfer Plan • Tech Transfer Team • UNC RENCE Partnership

Virtual Climate Data Server

IPCC Microservices

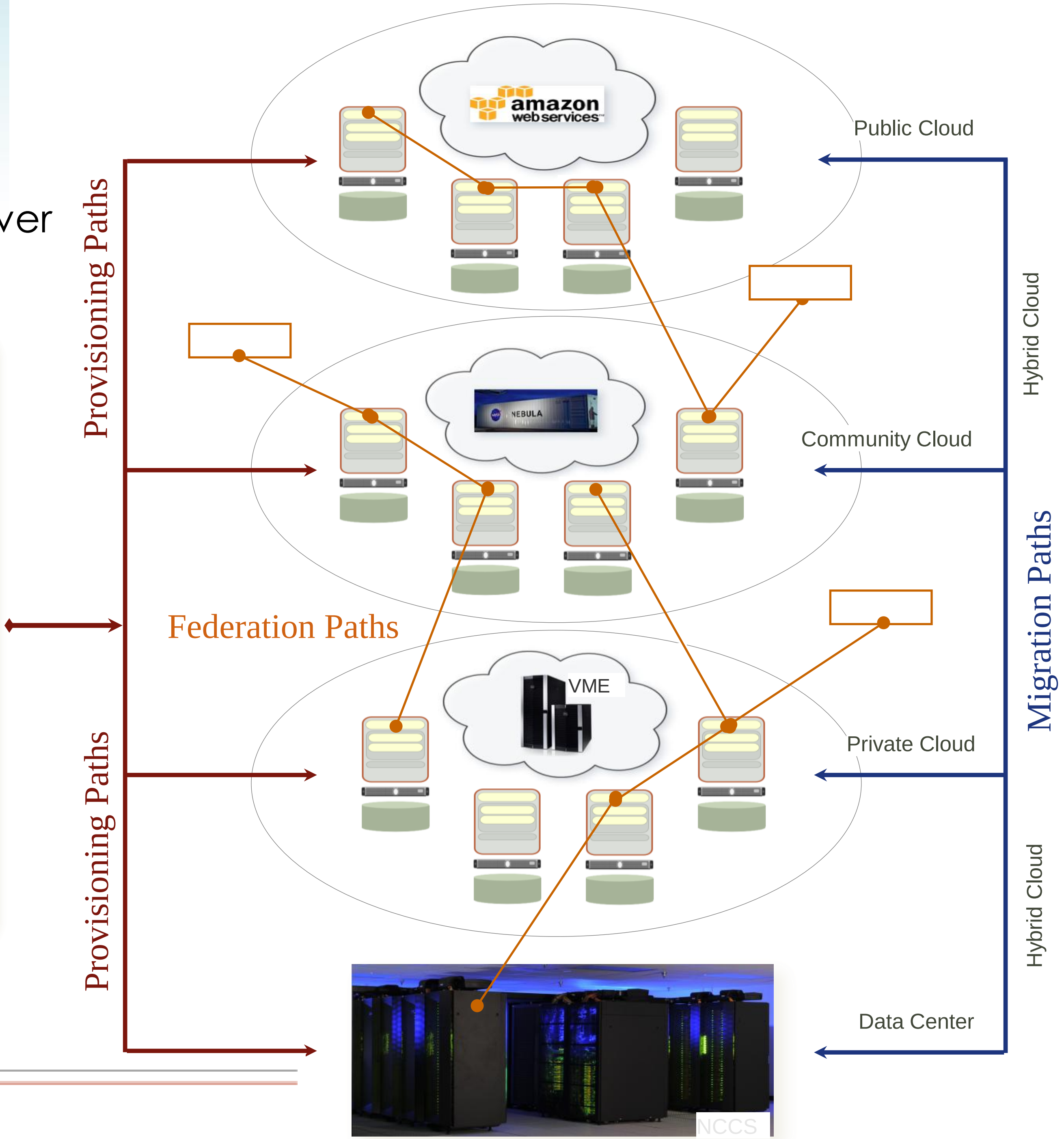
NetCDF iCAT

NetCDF Rules

iRODS 3.0 EE

SLES 11 SP 1

vCDS V0.9



Large-scale genomic research

The Sanger Institute

- Sequenced 1/3 of the human genome (largest single contributor).
- Active cancer, malaria, pathogen and genomic variation/human health studies.
- All data is made publicly available: websites, ftp, direct database access, programmatic APIs.
- Most-cited in the UK (Science Watch, 2008)
- Funded by WellcomeTrust: 2nd largest research charity in the world.
- More than 800 employees.
- Based in WellcomeTrust Genome Campus, Hinxton, Cambridge, UK. (share with EBI)

wellcome trust sanger institute

Large-scale genomic research

The Broad Institute at MIT and Harvard

- Launched in 2004: Eli and Edyth Broad, Harvard and affiliated hospitals, MIT

Gerstner Family Foundation gift supports groundbreaking effort in ADHD

Family gives $4.6M for program devoted to genomic studies of attention deficit hyperactivity disorder

First recurrent fusion gene identified in colorectal cancer genomes

Whole genome sequencing reveals unexpected genetic anomalies

Diving deep into mitochondrial diseases

MitoExome researchers search for the molecular basis of disease

Genetic variants for schizophrenia, bipolar disorder identified

International collaboration points to new connections between psychiatric diseases and biology of the brain

Data

Wherever your interests lie, data from across the Broad can be accessed from this starting point

Software

Scientists in the Broad community have developed many critical software tools for the analysis of increasingly large genome-related datasets, and they make these tools openly available to the scientific community

Careers

We are always looking for new team members to help us tackle important problems at the cutting edge of science

Our Approach

The Broad Institute is essentially an "experiment" in a new way of doing science

- Seek to:
  - Uncover the molecular basis of major inherited diseases
  - Unearth all the mutations that underlie different cancer types
  - Discover the molecular basis of major infectious diseases
  - Change therapeutic approaches
- Data contributed to NIH's NCBI (National Center for Biotechnology Information)
- Human Genome Project

Sequencing Data

- 2001: capillary sequencing - one sequencer produced about 115 kbp (kilobase pairs) per day (500-600 base pairs per reaction, 96 reactions per run, 2+ runs per day)
- Currently: 3000 Gbases/week (16 Gbytes/Gbase) – capillary sequencing (48TB/week, almost 10 TB/day)
- Small, affordable machines => many per lab (Sanger will have 20)
- New data analysis pipelines required for new sequencing techniques

Data management is becoming extremely complicated and important.

Data Growth and Management

- Sanger: Storage/compute doubles every 12 months; 2009: ~7 PB raw data
- Broad: generating 10TB of data/day
- Very good at data management for the sequencing pipeline:
  - Strict metadata controls (wet-lab doesn't begin until investigator fulfills data privacy and archiving info)
  - Automated analysis pipeline – bar codes, tracking, archiving
- But research begins *after* this point
- No idea what researchers have been doing with their data

Accidents waiting to happen...

Borrowed from Guy Coates, Sanger

- From: <*User A*> *(who left 12 months ago)*
- *I find the <project> directory is removed . The original directory is "/scratch/<User B> (who left 6 months ago)"*
- *..where is the directory?*
- *If this problem cannot be solved, I am afraid that <project> cannot be released.*
- Need a file tracking system for unstructured data !!
- Want an institute wide, standardised system.
  - Invest in people to maintain/develop it.

iRODS in Large-Scale Sequencing

Broad:

- Cancer genome analysis pipeline – archival data automatically stored in iRODS
- High-performance storage for a fee; iRODS archival storage in exchange for tagging the data
- Integration of OS authorization

iRODS in Large-Scale Sequencing

Sanger:

- Define a resource group, replicate data automatically to all resources in the group
- Delayed and periodic rules to check replication and repair holes in case of interruption (which always comes, sooner or later!)
- Authentication mechanisms – Kerberos, Shibboleth

Broad and Sanger will be working on a pilot iRODS federation.

iRODS Enterprise Version: iRODS-E

- RENCI's support for long-term iRODS sustainability
- Target new funding models – move beyond traditional public research funds
- 0.9 Release due out early next year
- RedHat Fedora model:
  - Community code (DICE) releases every 4 months
  - Enterprise code (RENCI) releases every 18 months
  - Product lines, service agreements based on iRODS-E

iRODS Enterprise Version: iRODS-E

- Properties of Enterprise Software:
  scalability, reliability, security, accessibility

- Achieved through:
  - Continuous Integration
  - Automated Testing
  - Code Hardening: static analysis, code review, refactoring, application of software engineering best practices
  - Documentation & Packaging: system-specific binary packages, user documentation, cookbook guides