

# Applying iRODS for Building an Integrated Data Archive (IDA)



# Talk Overview

- Introduction
  - CSC – IT Center for Science
  - Why we chose iRODS?
  - Current system setup
- Experiences in applying iRODS
  - Challenges and solutions
- Future plans

# CSC

- CSC - IT Center for Science Ltd is administered by the Finnish Ministry of Education and Culture.
- Provides Finland's most powerful supercomputing environment that researchers can use via the Funet network.
- Operates Haka identity federation (Shibboleth) of the Finnish universities and research institutions.



# Projects and Collaborations

- EUDAT
- Climate research projects
- National Digital Library, long-term preservation
- Finnish universities and research groups, data appraisal

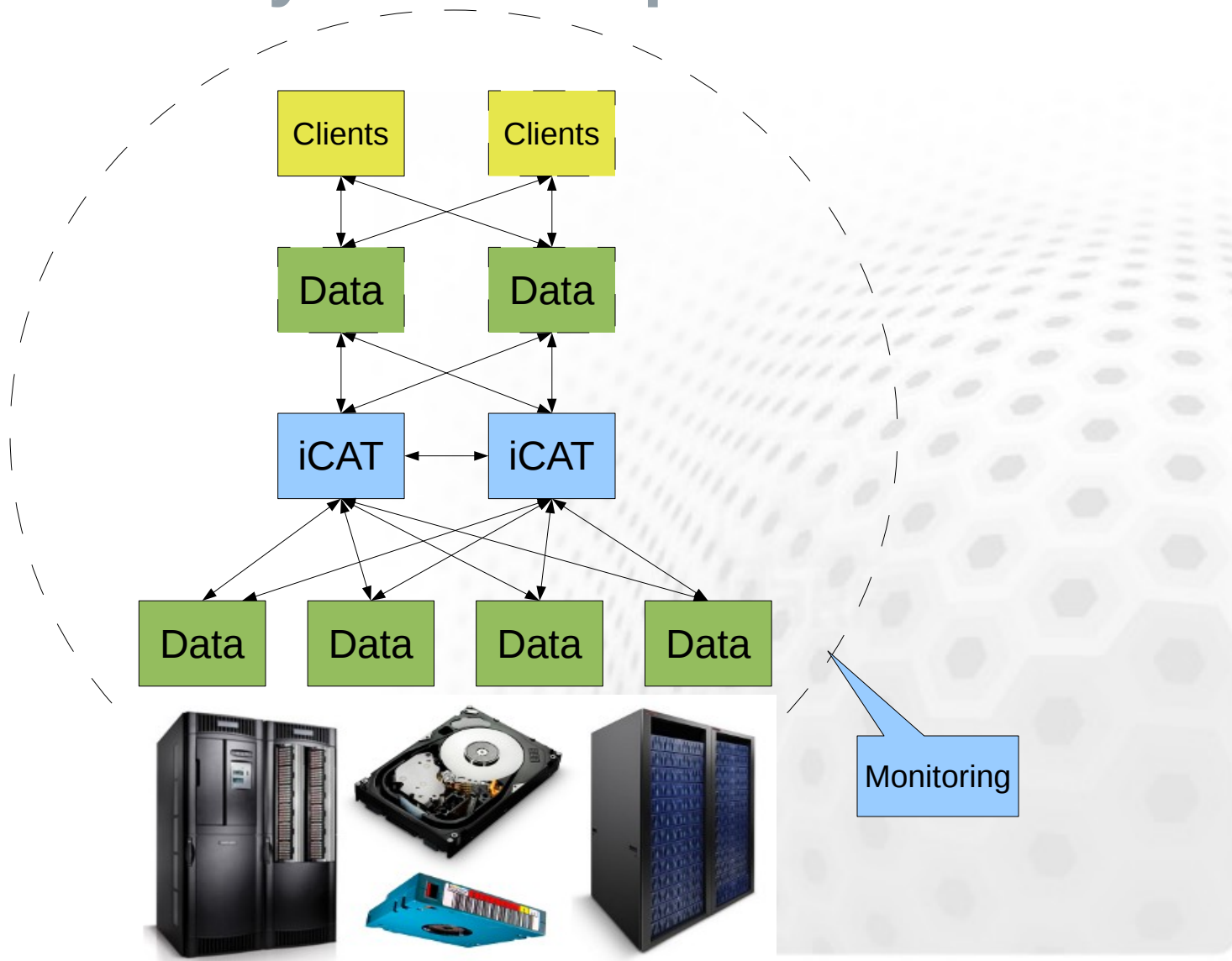


# Why we chose iRODS?

- Virtualization
  - Policies
  - Resources
- Storage management
  - Staging and migration support
- Metadata
- Distributed system
- Active iRODS community and open source



# System setup





# Experiences Overview

- Ingest workflow, metadata policy
- Shibboleth integration
- MSS interface to SAM-FS
- Securing iRODS protocol

## Ingest workflow

- Data is delivered as is: a lot of small data objects might be uploaded during a short time interval.
- There are computation intensive actions (i.e. virus scan for large files) included in the ingest workflow.
- The ingest workflow must be reliable.
- Thus, we have divided the workflow into several stages. The data object metadata includes the status of each stage.



# Shibboleth integration

- Existing Shibboleth integrations by King's College London (ASPiS project) and ARCS (Davis/WebDAV, Australia)
- The included one-time-password service requires user to be already authenticated.
- In addition, we need to map IdP asserted attribute (`eduPersonPrincipleName`) to local namespace.



Shibboleth.

## MSS interface to SAM-FS

- Due to security requirements, notify-pull interaction with a staging cache to SAM-FS
- A resource group including cache and compound resources
- Delayed rules
  - Archival status in metadata
  - Trimming of resources
- Bundling (iphybun or eq. microservice)

# Securing iRODS protocol

- Customers did not want to encrypt their data locally; data is delivered as is.
- Secure tunneling of the iRODS protocol over TLS/SSL
  - Caveat: requires a customer to install additional software, i.e. stunnel.

# Future plans

- Scientist's User Interface (Jargon)
- Moonshot





# Scientist's User Interface



Manage ▾



Jani Heikkinen (Sign Out)

Home Services Contact

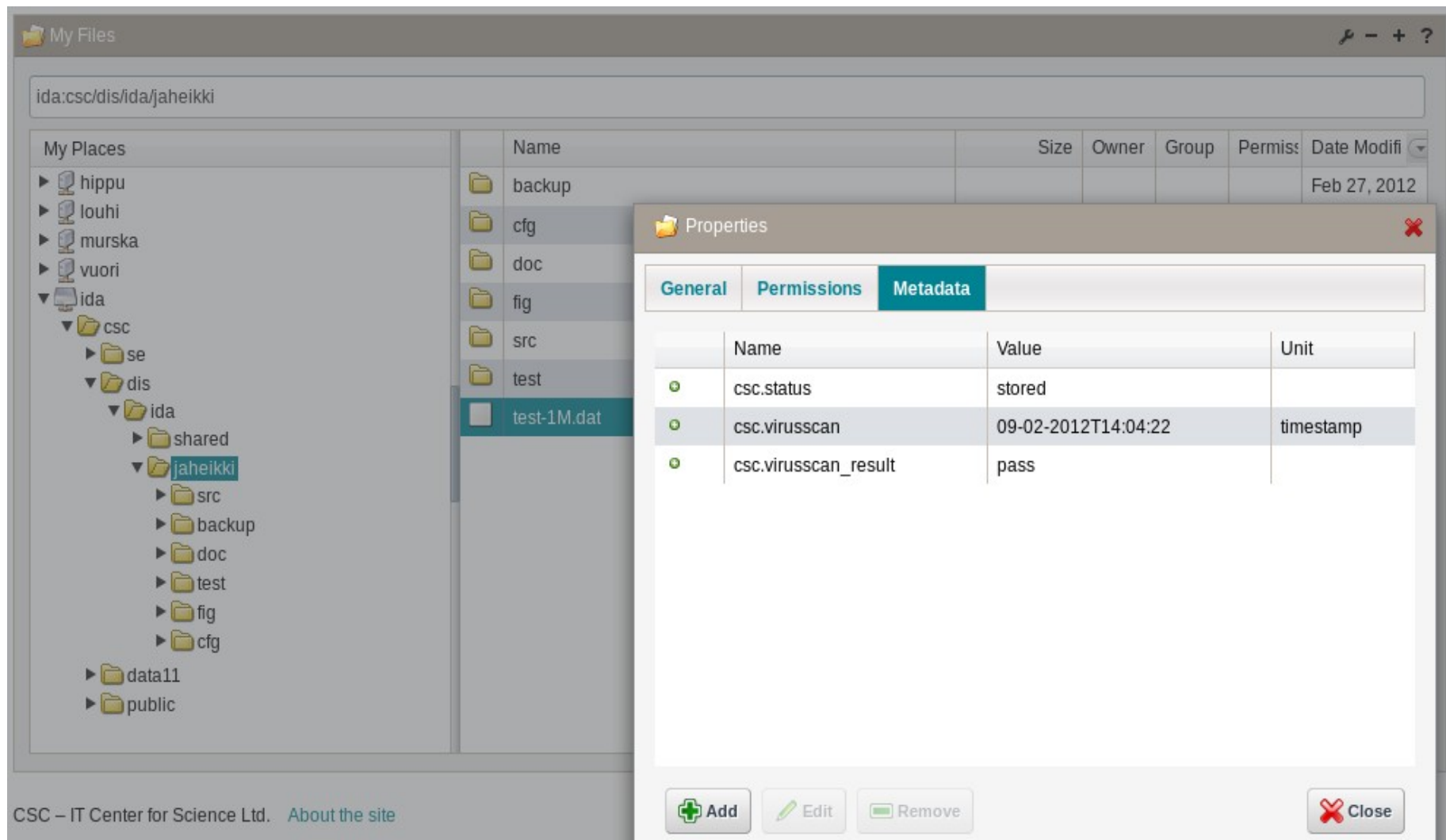
SUI > Services > My Files

My Files ⌘ - + ?

ida:csc

My Places	Name	Size	Owner	Group	Permiss	Date Modifi
▶ hippu	▶ data11					Jun 29, 2011
▶ louhi	▶ dis					Sep 14, 2011
▶ murska	▶ public					Apr 12, 2011
▶ vuori	▶ se					Jun 14, 2011
▼ ida						
▶ csc						
▶ se						
▶ dis						
▶ data11						
▶ public						

# Scientist's User Interface



The screenshot displays a web-based file management interface. On the left, a 'My Places' sidebar shows a tree view of folders, with 'ida' expanded to show sub-folders like 'csc', 'dis', and 'jaheikki'. The main area shows a file list with columns for Name, Size, Owner, Group, Permissions, and Date Modified. The file 'test-1M.dat' is selected. A 'Properties' dialog box is open over the file, showing the 'Metadata' tab with a table of metadata entries.

Name	Value	Unit
csc.status	stored	
csc.virusscan	09-02-2012T14:04:22	timestamp
csc.virusscan_result	pass	

At the bottom of the interface, there are buttons for '+ Add', 'Edit', 'Remove', and 'Close'. The footer text reads 'CSC - IT Center for Science Ltd. [About the site](#)'.

## Moonshot integration

- “Moonshot project aims at developing a general approach for associating a federated identity with arbitrary Internet protocols.”
- iRODS GSS-API
- Required infrastructure: RADIUS/Diameter, Shibboleth components, ...
- Our goal is to learn more about the deployment: how organizations can take up the technology, what value it adds, etc.