Applying iRODS for Building an Integrated Data Archive (IDA)

iRODS User Meeting 2012, Tucson

jani.heikkinen@csc.fi

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.

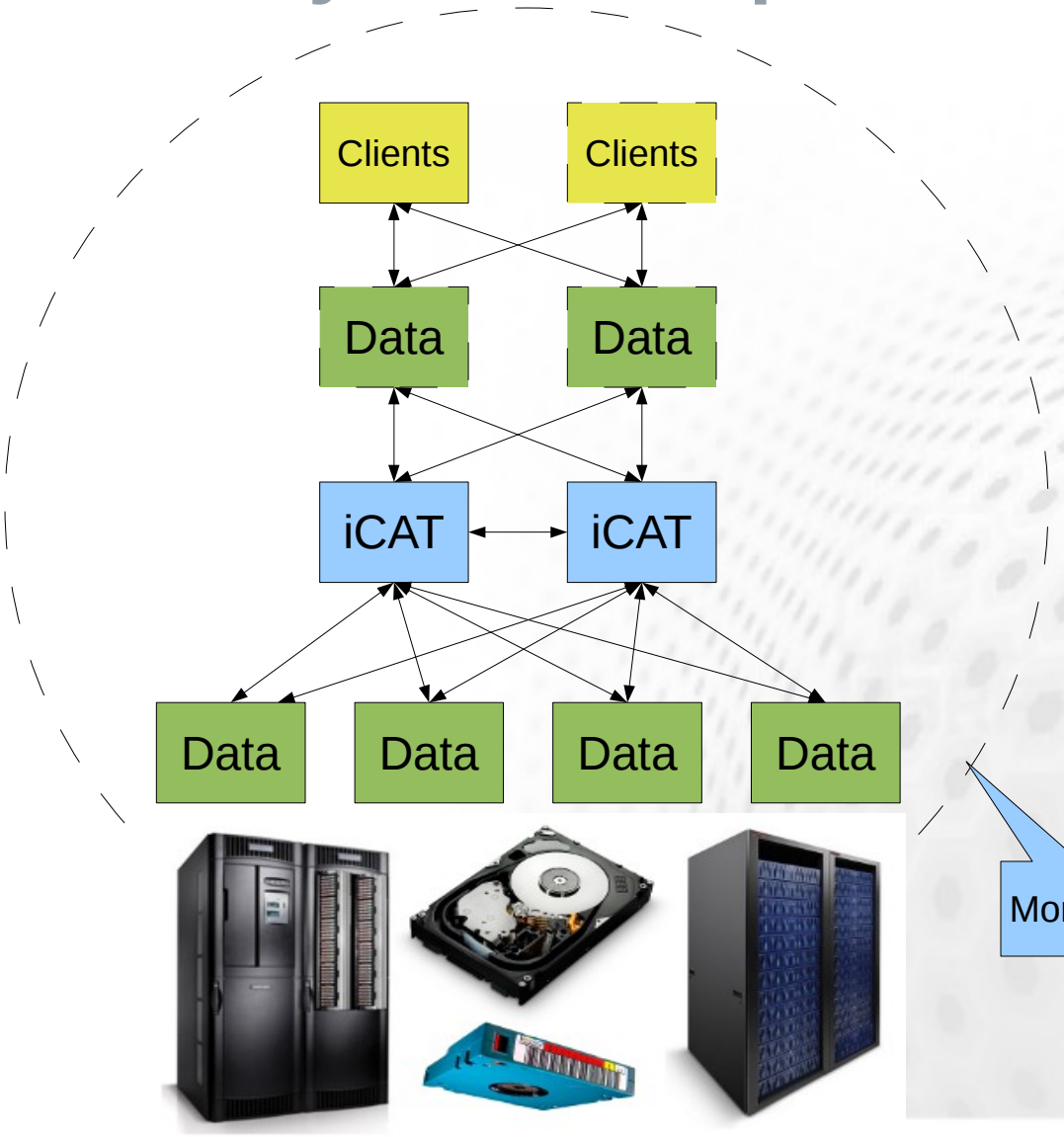
Projects and Collaborations

- 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.

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)

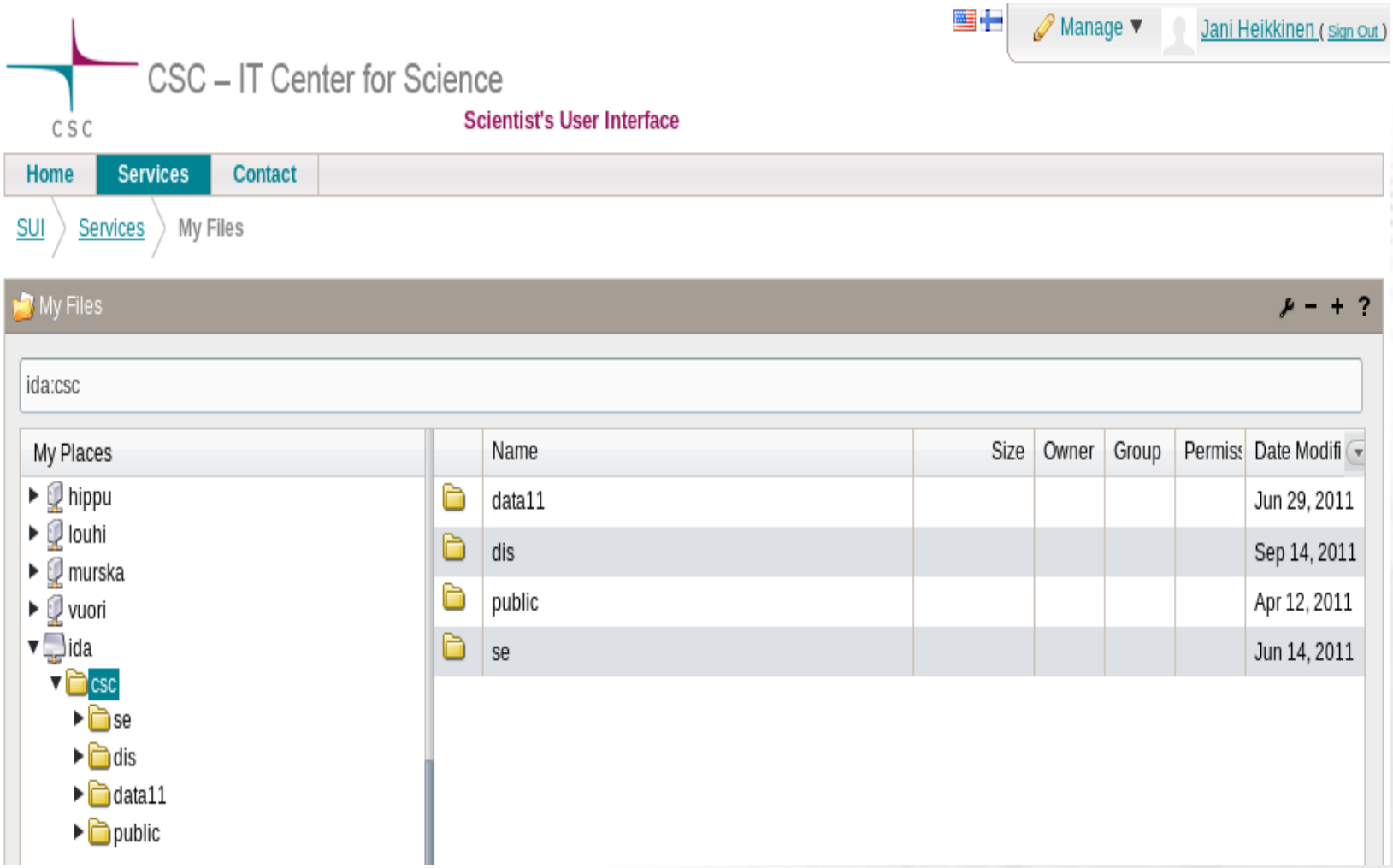
Securing iRODS protocol

- 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.

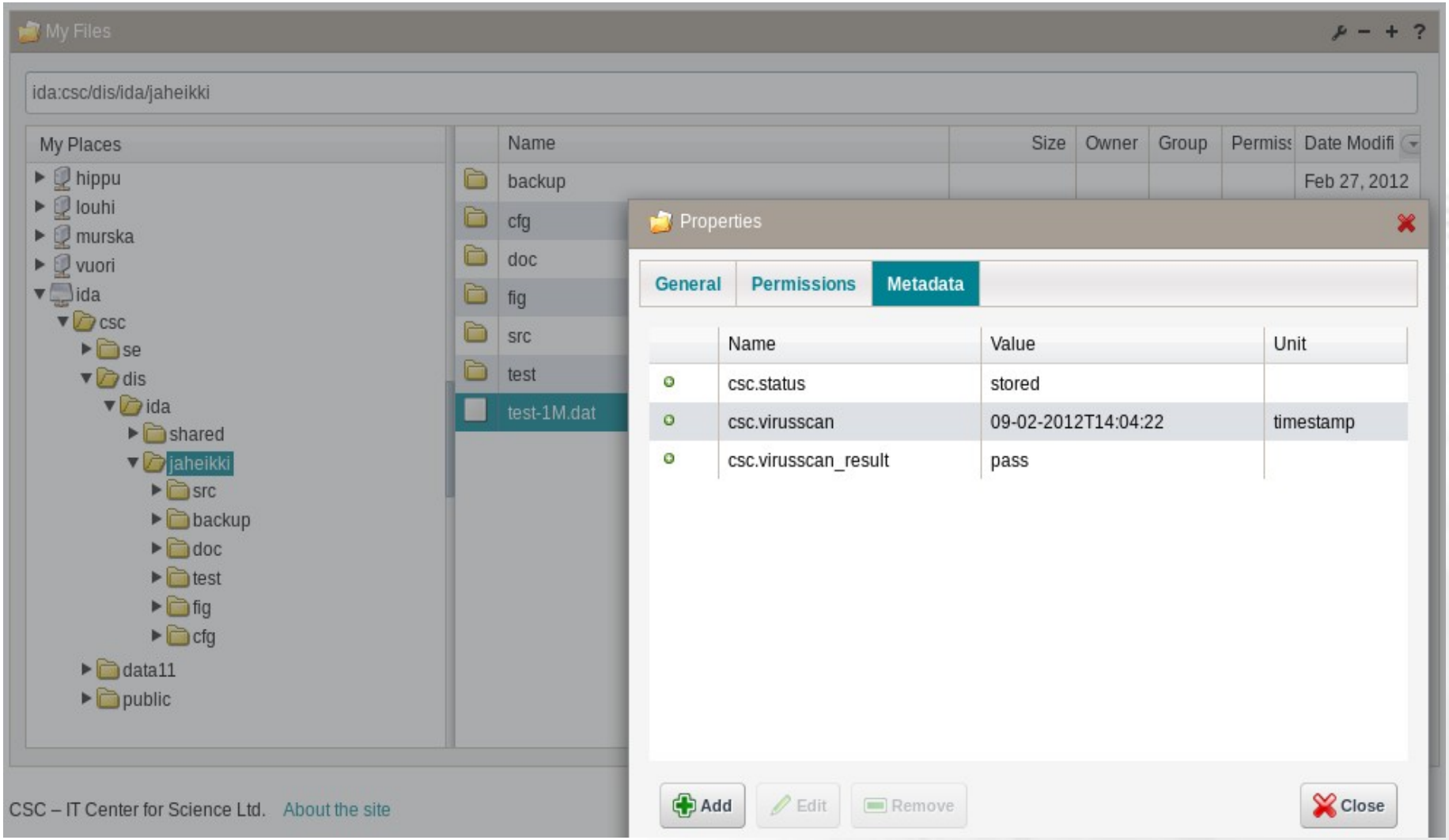
Future plans

- Scientist's User Interface (Jargon)
- Moonshot

Scientist's User Interface



Scientist's User Interface



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.