

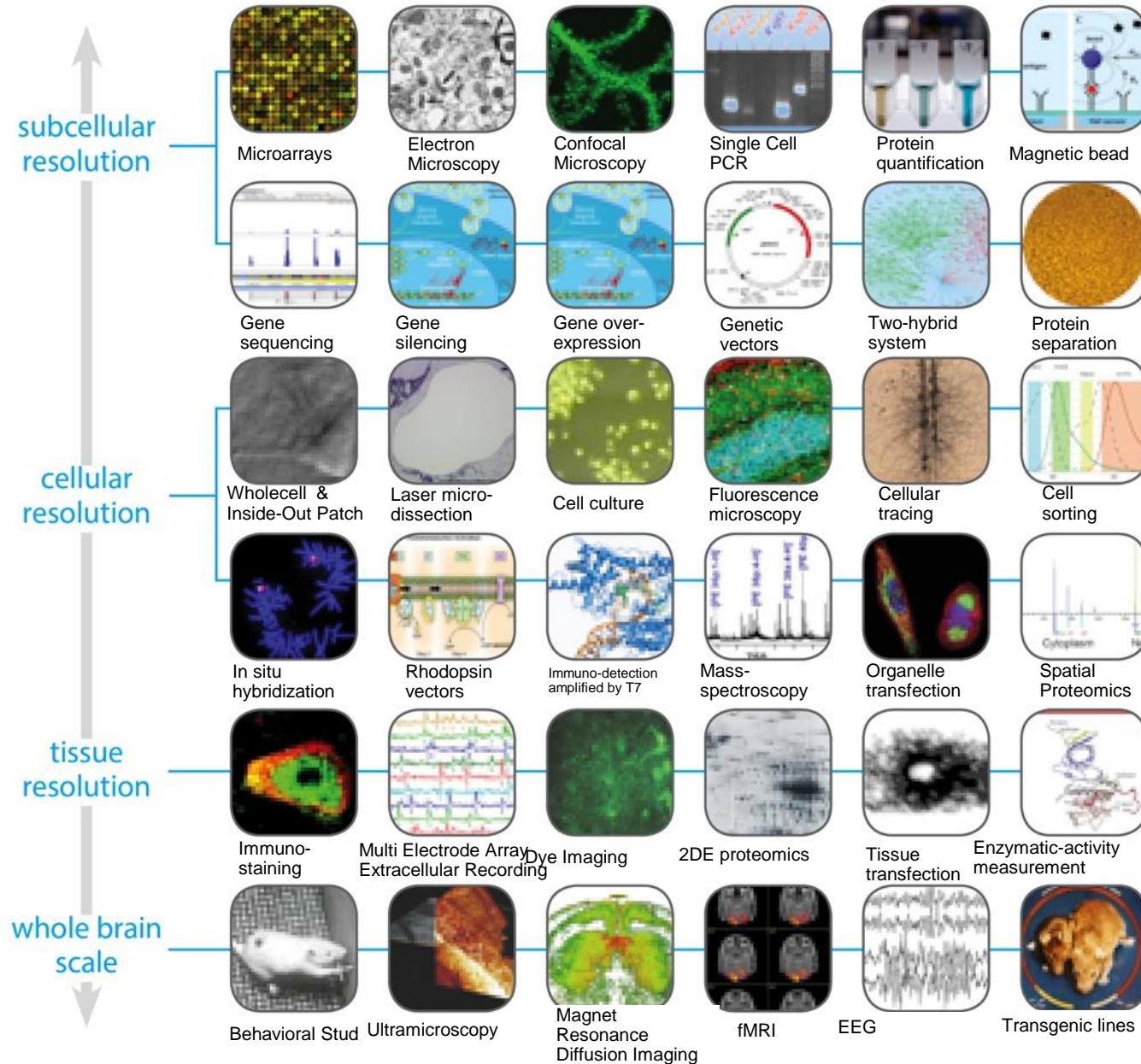
# IDS – The INCF DataSpace

Raphael Ritz, Scientific Officer  
International Neuroinformatics Coordinating Facility  
Stockholm, Sweden

[raphael.ritz@incf.org](mailto:raphael.ritz@incf.org)

iRODS User Group Meeting, February 28, 2013, Garching, Germany

# Multiomic Neuroscience Data



to replicate experiments

to visualize

to ask new questions

to simulate

to analyze

How do we bring all this data together?

to model

to search

to share

to publish

to teach

# The Birth of INCF

- The Global Science Forum of OECD realized the need for a concerted action for developing Neuroinformatics on the international level
- 2005 INCF plans endorsed by the ministers of research of OECD
- August 1st 2005 INCF formed with 7 members including Japan and the US

# The mission of INCF

- Coordinate and foster international activities in neuroinformatics
- Contribute to development and maintenance of database and computational infrastructure and support mechanisms for neuroscience applications
- Enable access to all freely accessible data and analysis resources for human brain research to the international research community
- Develop mechanisms for the seamless flow of information and knowledge between academia, private enterprises and the publication industry

# In general data sharing is difficult

“Where do I put my data to share it?”

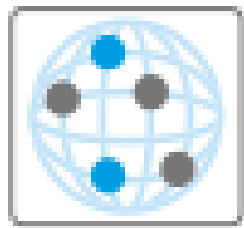
“How can I share my data with you  
(and only you)?”

“Where can I backup my data?”

“Where can I look for shared data?”

# How can we make it easier?

- Let's make data sharing as simple as possible - like a Dropbox for Scientists
- Drag and drop any type of data, text, images
- Don't worry about metadata (yet)



# DATASPACE

[ids.incf.net](http://ids.incf.net)





# DATASPACE

[dataspace.incfn.org](http://dataspace.incfn.org)

Do you  
want to...?

... discover  
neuroscience  
data?



... make your  
data visible  
globally?



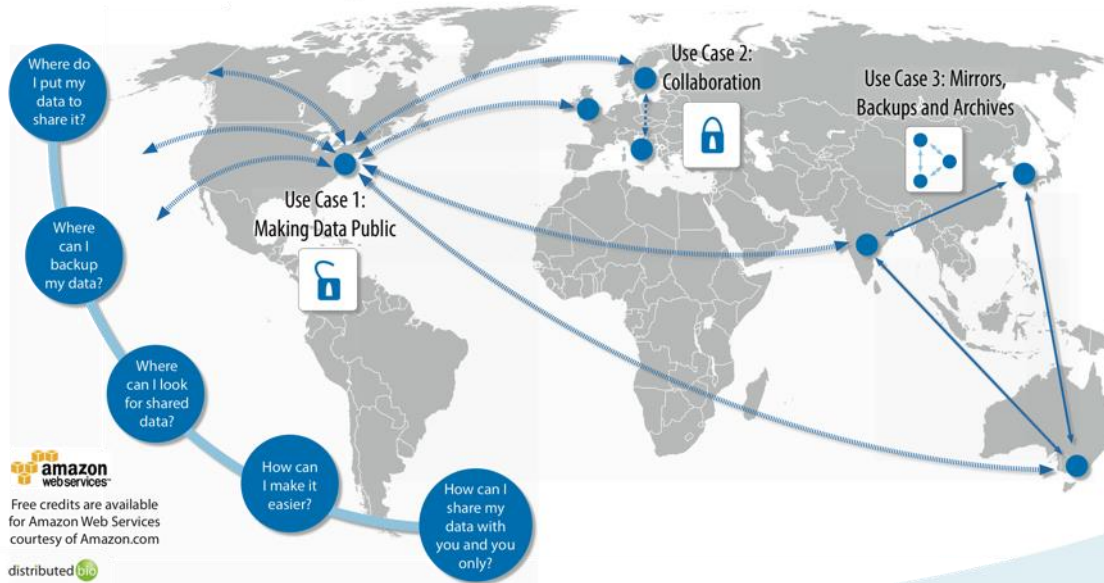
... make  
your data  
visible to  
collaborators?



... share your  
existing  
large  
dataset?



Connect  
today!



## Features...

- Access diverse data repositories from around the world through a single resource
- Browse and access data using different user interfaces
  - Web
  - File navigator
  - Command line
- Upload and download data worldwide
- Set/get arbitrary metadata for files and folders
- Search metadata
- Manage large data
- Keep directories synchronized
- Create temporary public or private links to share data

connect share win

Are you  
producing  
valuable data?



Are you  
producing  
reusable data?

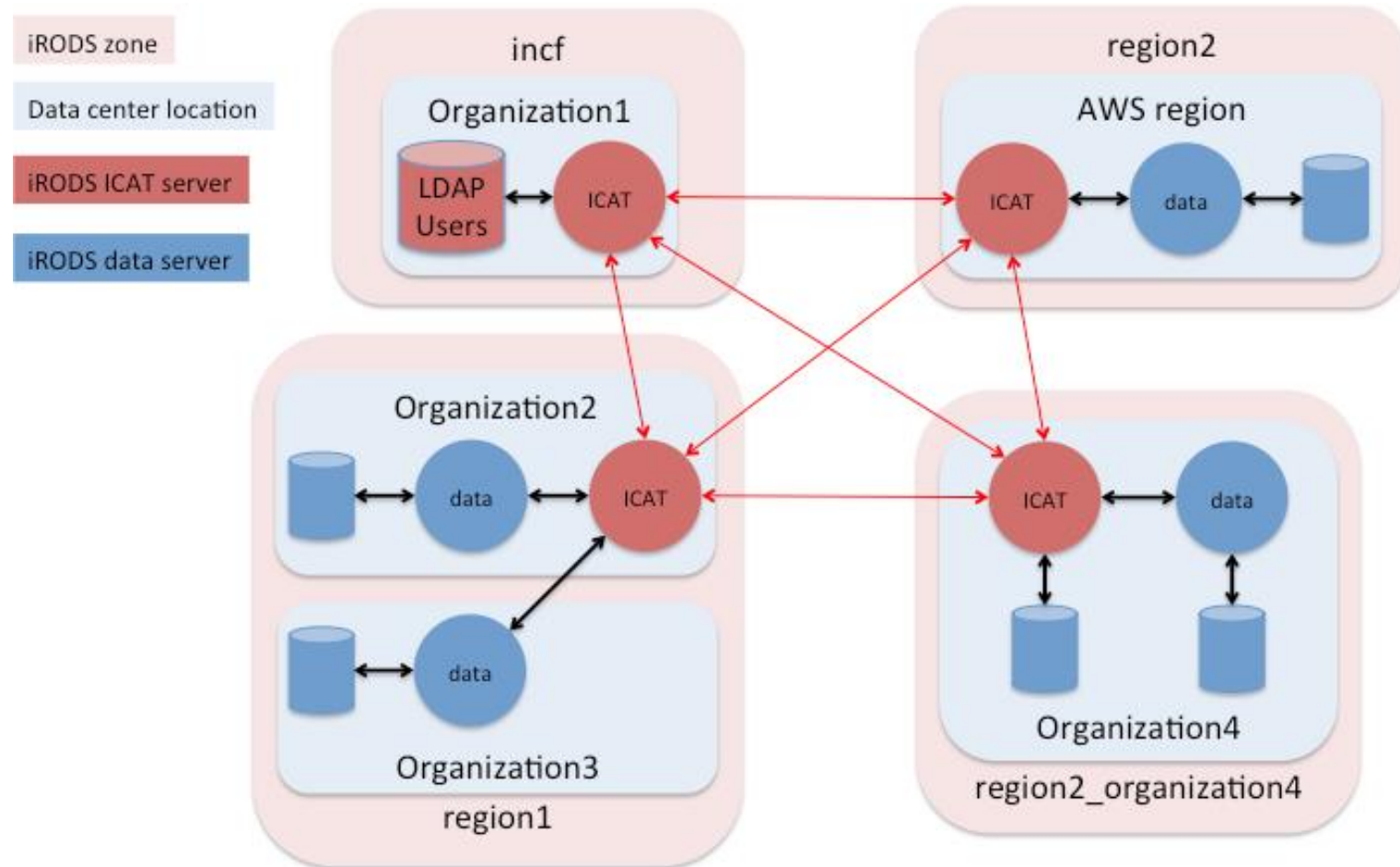


## data ShareAward

[incfn.org/datashareaward](http://incfn.org/datashareaward)

**incfn**

# INCF Data Space (IDS) - Architecture



# Deployment

- Central servers in the Amazon Cloud (EC2)
  - Replicated across 4 availability zones
  - Master in Europe
  - Slaves in US-East, US-West, AP-NE
- Community contributed data and zone servers
  - Debian packages (RPMs coming)
  - EC2: Region-specific cloud formation templates
- IDS Tools: utilities to setup and maintain servers

# Information Architecture

- Users have home folders in the INCF zone backed by INCF-managed resource servers (quotas enforced)
- Contributed data servers are hooked up at
  - `/incf/resources/<reverse domain name>`
- Rules define and enforce which resource receives uploads based on location in namespace

# Web Interface: ids.incf.net

← → ↻ <https://ids-use.incf.net/idrop-web/#absPath=/incf/resources/org.incf/public/WaxholmMouseAtlas&browseOptionVal=browse> ☆ ☰

incf INCF main website

**incf** **DATASPACE**




User Name:ritz Zone:incf Resource:  Logout Show/Hide Side Bar

Search as a **file** or  Open the path in the tree

**Browse** **Search**

/ incf / resources / org.incf / public / WaxholmMouseAtlas

File » View » Upload and Download » Tools » Apply an action to all selected items »

  
incf  
└─ home  
└─ resources  
    └─ com.distributedbio  
    └─ edu.washington.ibic  
    └─ org.incf  
        └─ private  
        └─ public  
            └─ WaxholmMouseAtlas  
            └─ INCF Dataspace Architecture v1.3  
            └─ INCF Dataspace Policies v1.3

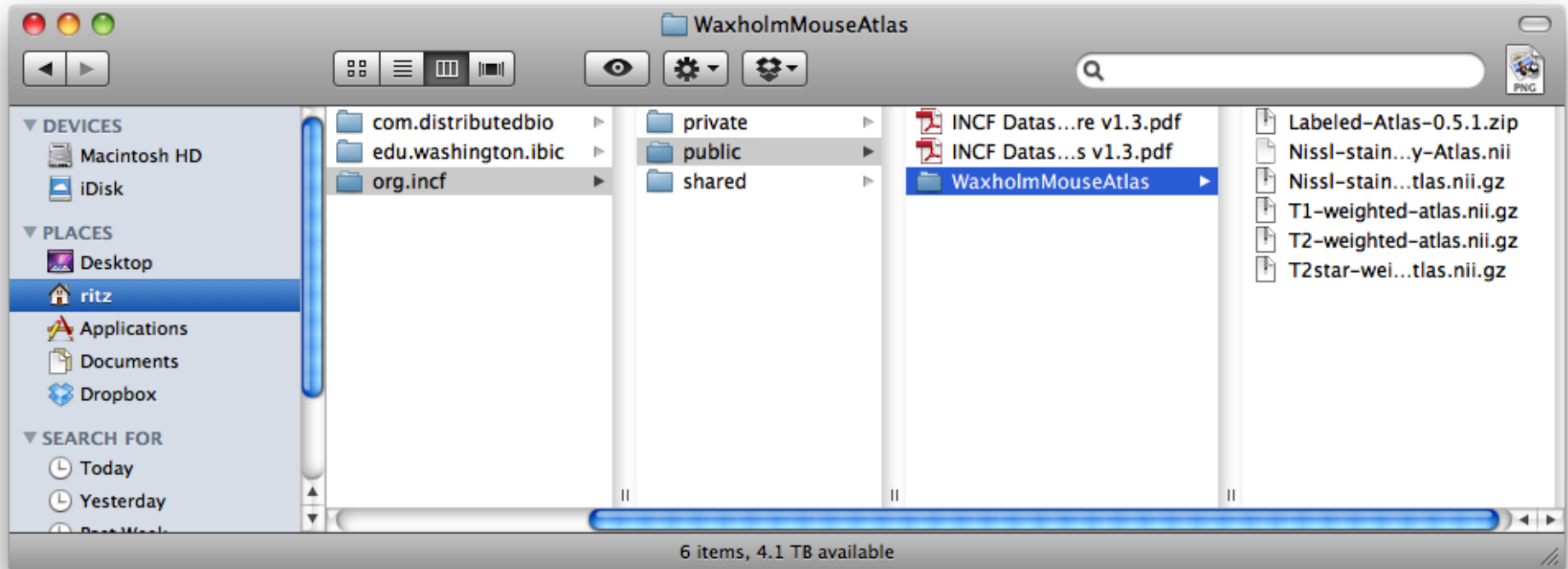
	Name	Type	Modified date	Length
<input type="checkbox"/>	<b>Labeled-Atlas-0.5.1.zip</b>	DATA_OBJECT	Sun Oct 14 18:58:44 UTC 2012	1 GB
<input type="checkbox"/>	<b>Nissl-stained-Optical-Histology-Atlas.nii</b>	DATA_OBJECT	Mon Oct 15 15:01:57 UTC 2012	512 MB
<input type="checkbox"/>	<b>Nissl-stained-Optical-Histology-Atlas.nii.gz</b>	DATA_OBJECT	Sun Oct 14 18:37:07 UTC 2012	41 MB
<input type="checkbox"/>	<b>T1-weighted-atlas.nii.gz</b>	DATA_OBJECT	Sun Oct 14 18:03:24 UTC 2012	432 MB
<input type="checkbox"/>	<b>T2-weighted-atlas.nii.gz</b>	DATA_OBJECT	Sun Oct 14 18:28:41 UTC 2012	325 MB
<input type="checkbox"/>	<b>T2star-weighted-Atlas.nii.gz</b>	DATA_OBJECT	Sun Oct 14 18:05:08 UTC 2012	385 MB

Showing 1 to 6 of 6 entries

# Command Line Client: icommands

```
Terminal — bash — 70x18
Last login: Tue Oct 23 07:57:27 on ttys000
Raphael-Ritzs-MacBook-Pro:~ ritz$ ipwd
/incf/home/ritz
Raphael-Ritzs-MacBook-Pro:~ ritz$ icd /incf/resources/org.incf/public
Raphael-Ritzs-MacBook-Pro:~ ritz$ ils WaxholmMouseAtlas
/incf/resources/org.incf/public/WaxholmMouseAtlas:
  Labeled-Atlas-0.5.1.zip
  Nissl-stained-Optical-Histology-Atlas.nii
  Nissl-stained-Optical-Histology-Atlas.nii.gz
  T1-weighted-atlas.nii.gz
  T2-weighted-atlas.nii.gz
  T2star-weighted-Atlas.nii.gz
Raphael-Ritzs-MacBook-Pro:~ ritz$
```

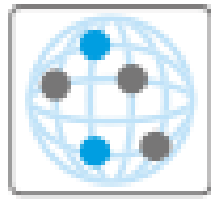
# Desktop Integration: irodsFuse





- INCF central authentication
- User defined access control (Private, Public, Group)
- Policy based group data access (e.g. data use agreement)
- Standardized navigation structure and policies
- Globally distributed zones - distributed data storage costs





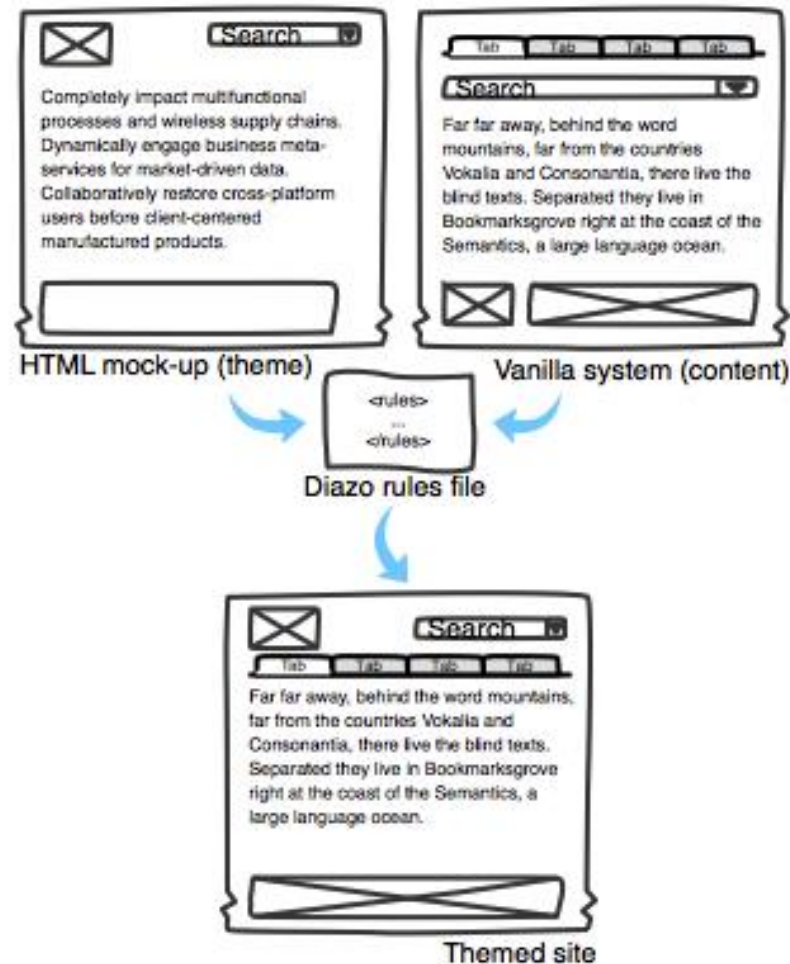
# DATASPACE

- Built existing technology – iRODS
- Scales with the Amazon Cloud
- Supports data replication across the federation
- Planning on federated search using NIF portal ([neuinfo.org](http://neuinfo.org))
- Provides strong data management foundation for future developments (arbitrary metadata, provenance, replication, archival, etc)



- Things we needed to add:
  - PAM support to authenticate against the INCF LDAP
  - Storage admin user to avoid the propagation of rodsadmins
- Thanks to Chris Smith, Wayne Schroeder and Mike Conway for the implementation.

# Theming the web ui: diazo.org



# Growing the Federation

- Challenges
  - People already have “some systems” – need to fit existing environments
  - EC2 is hard to pay for - and not necessarily cheaper than a university environment
  - Integrate at application rather than file level
- EUDAT
  - Simple Storage
  - Safe Replication
  - Persistent Identifiers

# Further Information

- Web access to the data space: <https://ids.incf.net>
- High level information: <http://dataspace.incf.org>
- Tools and clients: <http://github.com/INCF/ids-tools/wiki>
- Developers corner:
  - <http://dev.incf.org/trac/infrastructure>
  - <http://github.com/INCF/ids-tools>
- Contact: [ids-admin@incf.org](mailto:ids-admin@incf.org)

# Documentation

- For end users: video tutorials
  - <http://www.youtube.com/user/INCForg>
- Design documents
  - <http://dev.incf.org/trac/infrastructure/wiki>
- For administrators: data&zone servers
  - <http://github.com/INCF/ids-tools/wiki>
- Background reading: a workshop report
  - <http://www.incf.org/programs/workshops/scientific-workshops/ci-1>

# Contributors

- Sean Hill
- Chris Smith
- Sina Khaknezhad
- Ylva Lillberg
- Beatriz Martin
- Mathew Abrams
- EUDAT
  - Johannes Reetz
  - Dejan Vitlacil

incf@

Google

Contact info: [gsoc@incf.org](mailto:gsoc@incf.org)

Web:  
[www.incf.org/gsoc](http://www.incf.org/gsoc)

