

Development of a high throughput gene, environment and epigenetics database and analysis system for international ALS research

@ ENCALS 2017

Alfredo Iacoangeli

Al-Chalabi Group (Clinical Neuroscience)

Dobson Group (Health Informatics)

alfredo.iacoangeli@kcl.ac.uk

Issues

- Big Data

- how to store it
- how to manage it
- how to analyse it

- Collaboration

- ownership heterogeneity
- data sharing

- Audience

- accessibility
- impact

iRODS



Global Alliance
for Genomics & Health



docker



Bitbucket



github
SOCIAL CODING



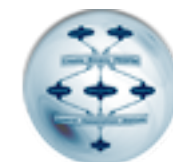
tranSMART
v1.2



Galaxy
PROJECT



Pachyderm

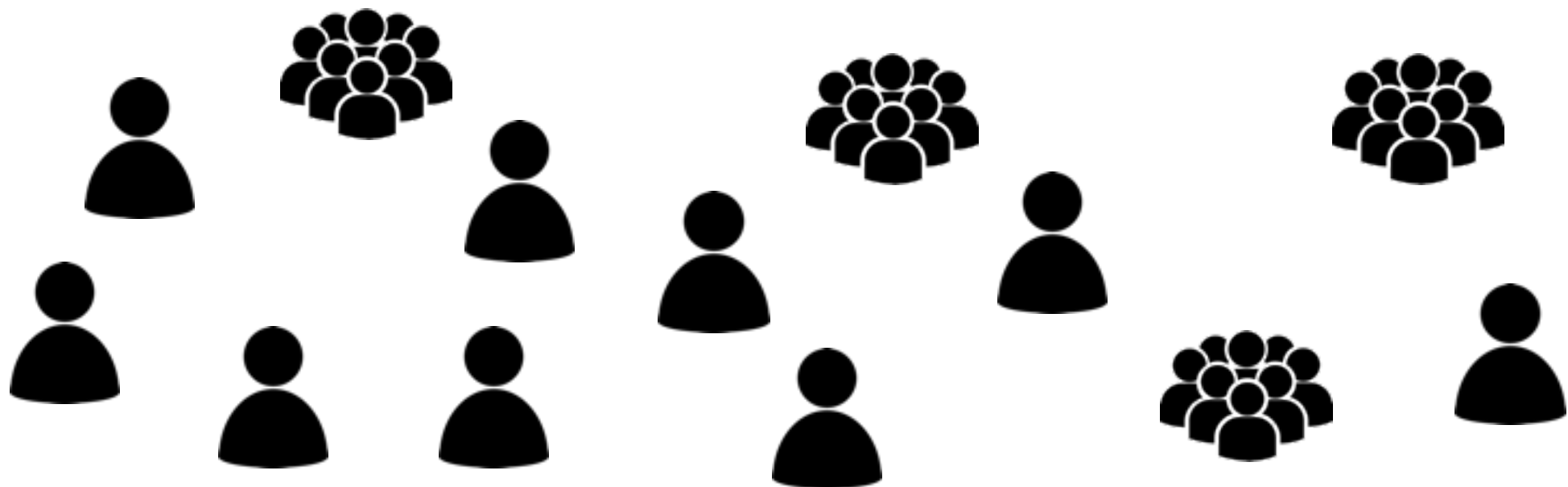


WINGS

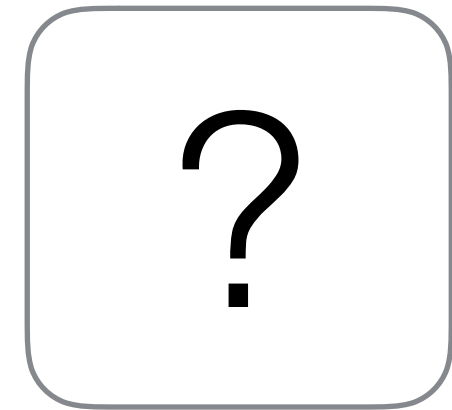
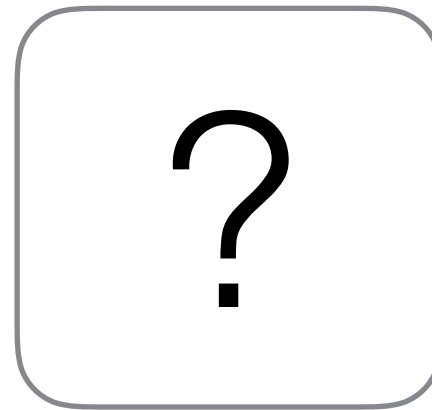
Data Management



Data Virtualisation Layer



Data Management



Data Virtualisation Layer



iRODS

iRODS is an open source software for:

- Working with data distributed across storage technologies
- Annotating and searching data with rich metadata
- Implementing access control, auditing, preservation, organisation, and data movement policies
- Providing a single interface to share data between organisations



Universiteit Utrecht



How would we interact with all of this?

- through any iRODS zone of the network
- with the terminal command line: `icommands`
- web-browser: search, download/upload, write rules, add metadata, more to come...

```

Terminal — bash — 101x36

-ChangeSettings ## Change system settings
-ObserveOnly ## Modify ControlObserve option to allow Observe mode only

-wait number ## Specify "wait" wait numerically instead (advanced)

-allowAccessFor ## Specify the Remote Management access mode
-allUsers ## Grant access to all local users
-specifiedUsers ## Only grant access to users with privileges

-computerInfo ## Specify all four computer info fields (default for each is empty)
-set1 -1 <text>
-set2 -2 <text>
-set3 -3 <text>
-set4 -4 <text>

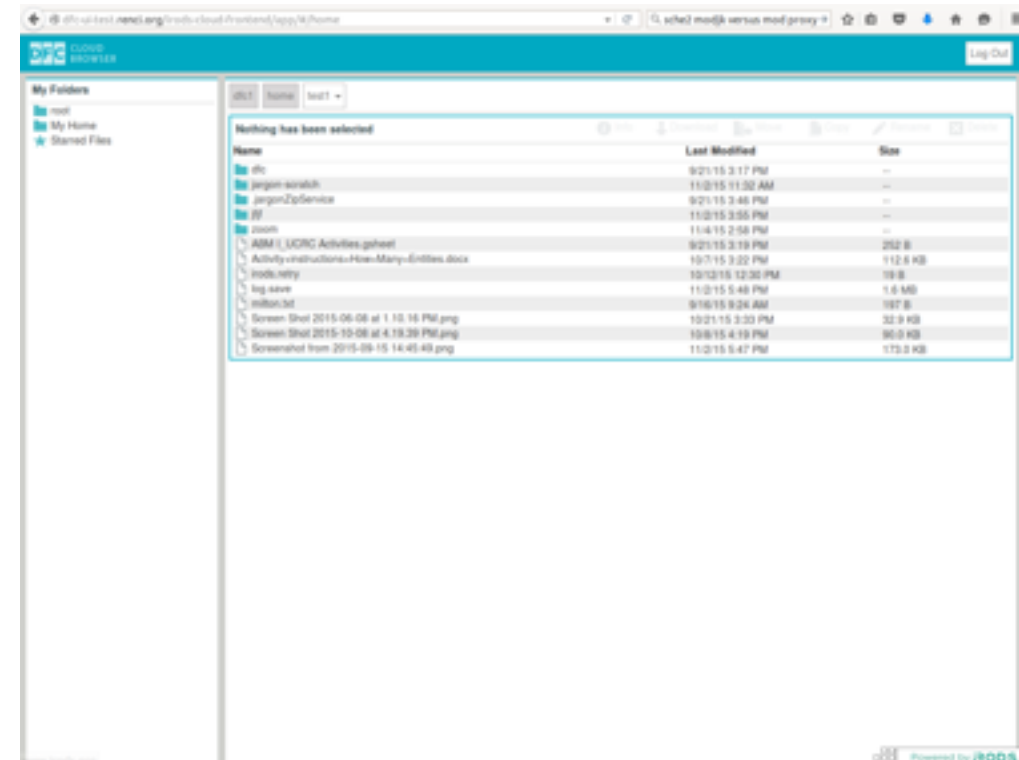
-clientOpts ## Allow specification of several opts.
-setMenuExtra -menuextra yesno ## Set whether menu extra appears in menu bar
-setDirLogins -dirlogins yesno ## Set whether directory logins are allowed
-setReqPerm -reqperm yesno ## Allow VNC guests to request permission
-setVncLegacy -vnclegacy yesno ## Allow VNC legacy password mode
-setVncPw -vncpw yesno ## Set VNC legacy PW
-setVncWeb -vncweb yesno ## Allow incoming VNC requests over IP

-stop ## Stop the agent and/or console program (N/A if targetdisk is not /)

-restart ## Enable the "restart" options: (N/A if targetdisk is not /)

-agent ## Restart the ARD agent and helper
-console ## Restart the console application
-menu ## Restart the menu extra

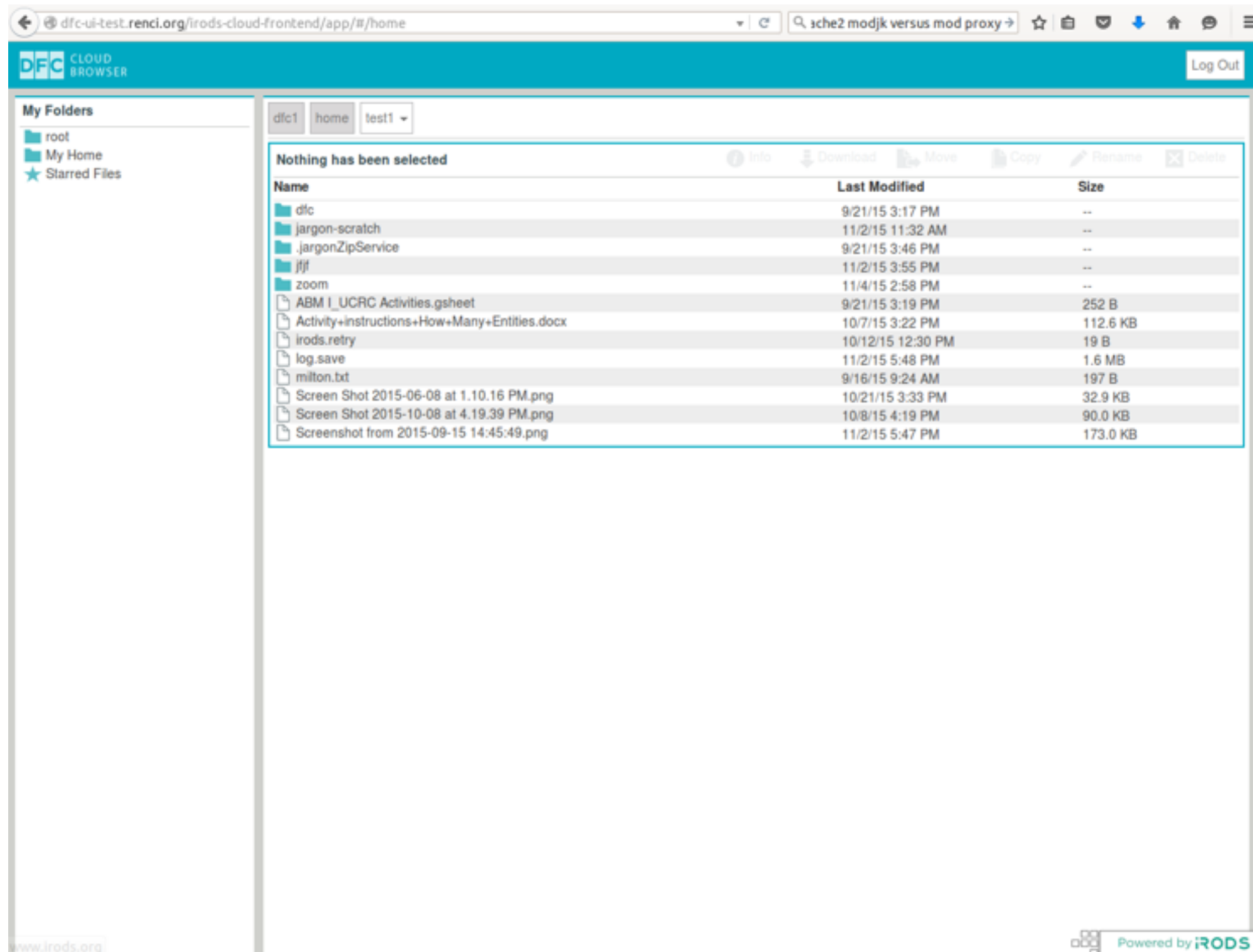
-targetDisk ## Disk on which to operate, specified as a mountpoint in
## the current filesystem. Defaults to the current boot volume: "/".
## NOTE: Disables the -restart options (does not affect currently
## running processes).
  
```



iRODS

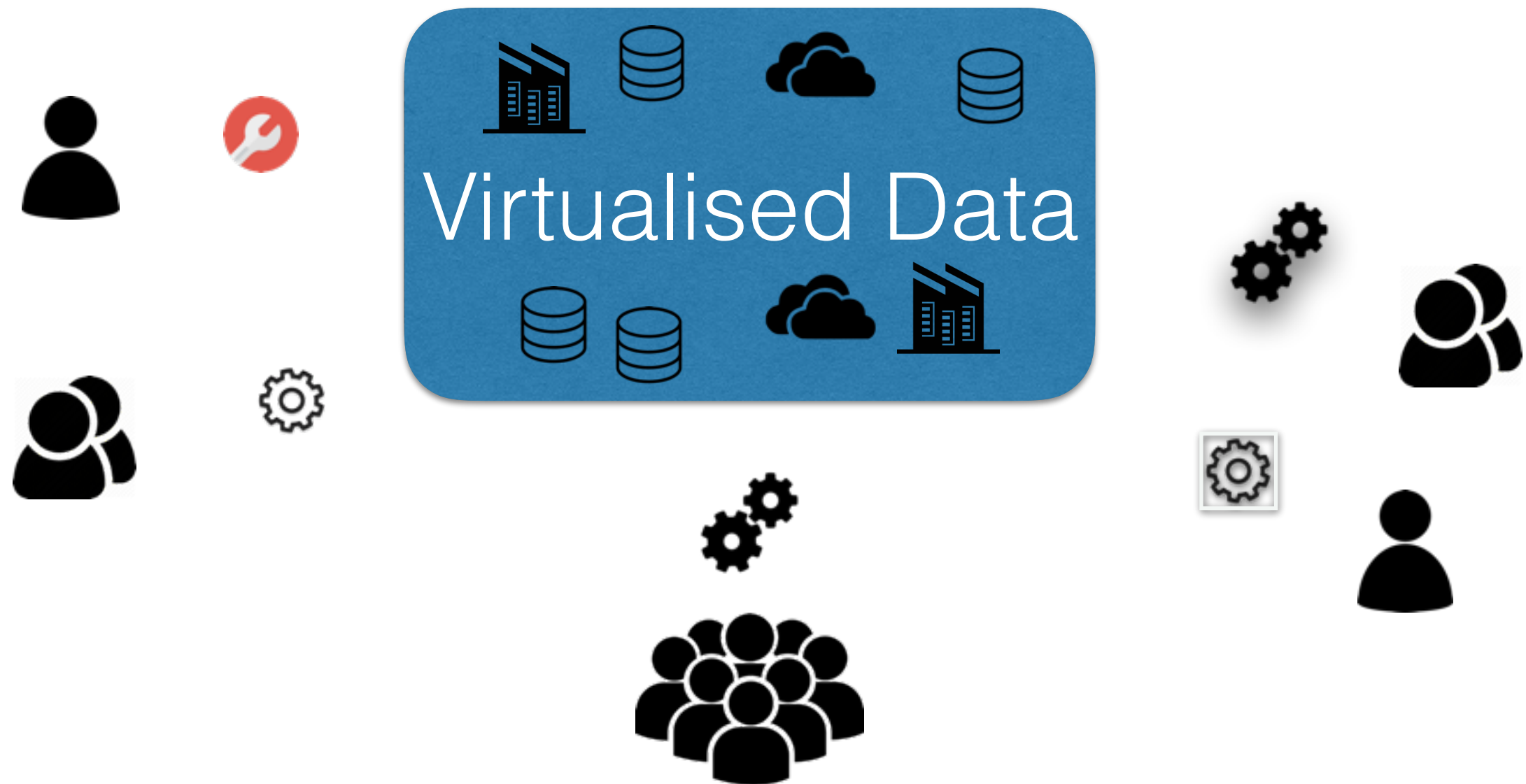
Web Browser

- Creation of new files
- File Editing
- Rule Execution
- Main Navigation Bar
- Metadata Search
- Recent Query Collections
- Docker deployment

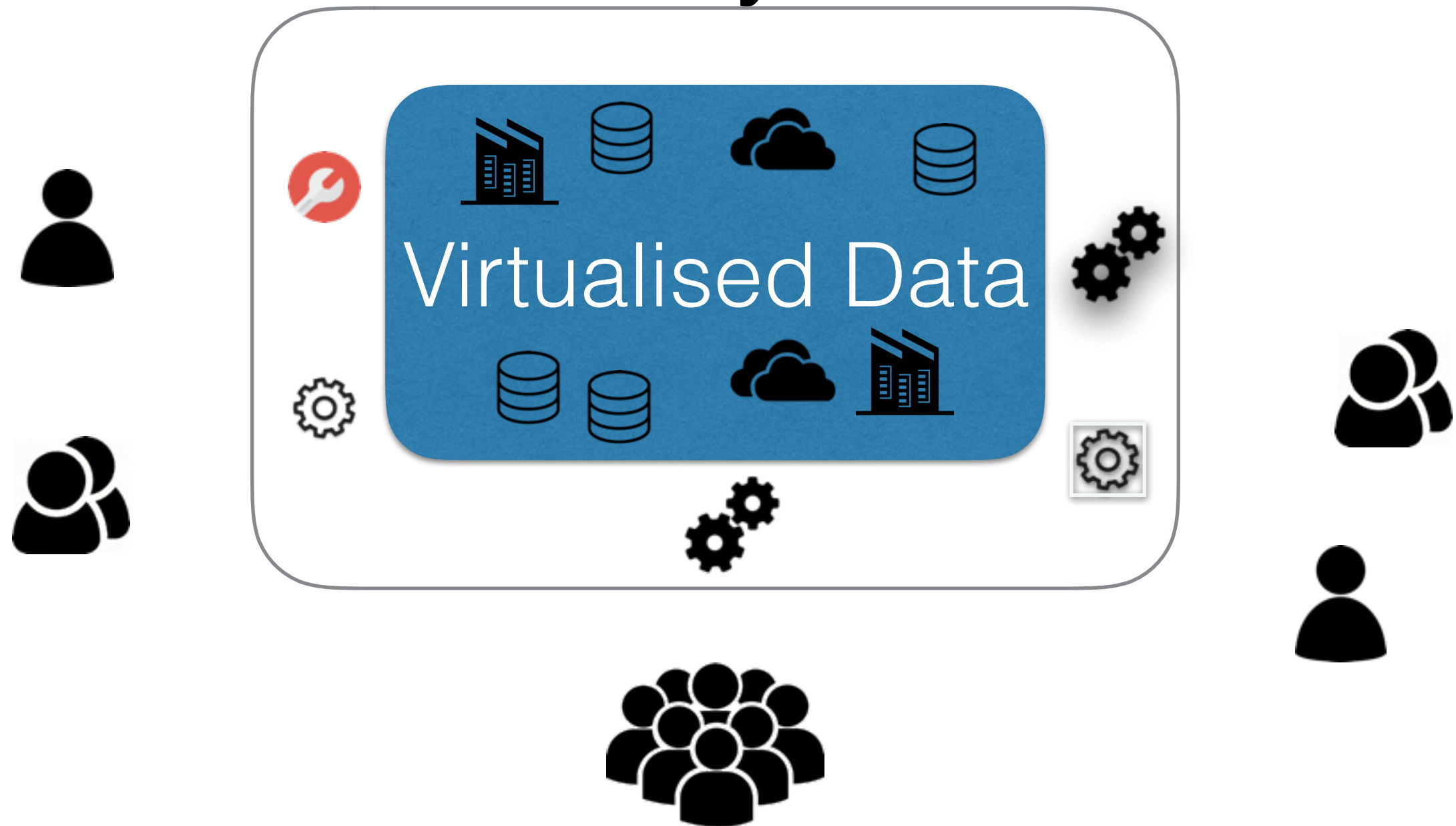


<http://52.202.127.9:8552/irods-cloud-frontend/>

Data Accessibility and Analysis



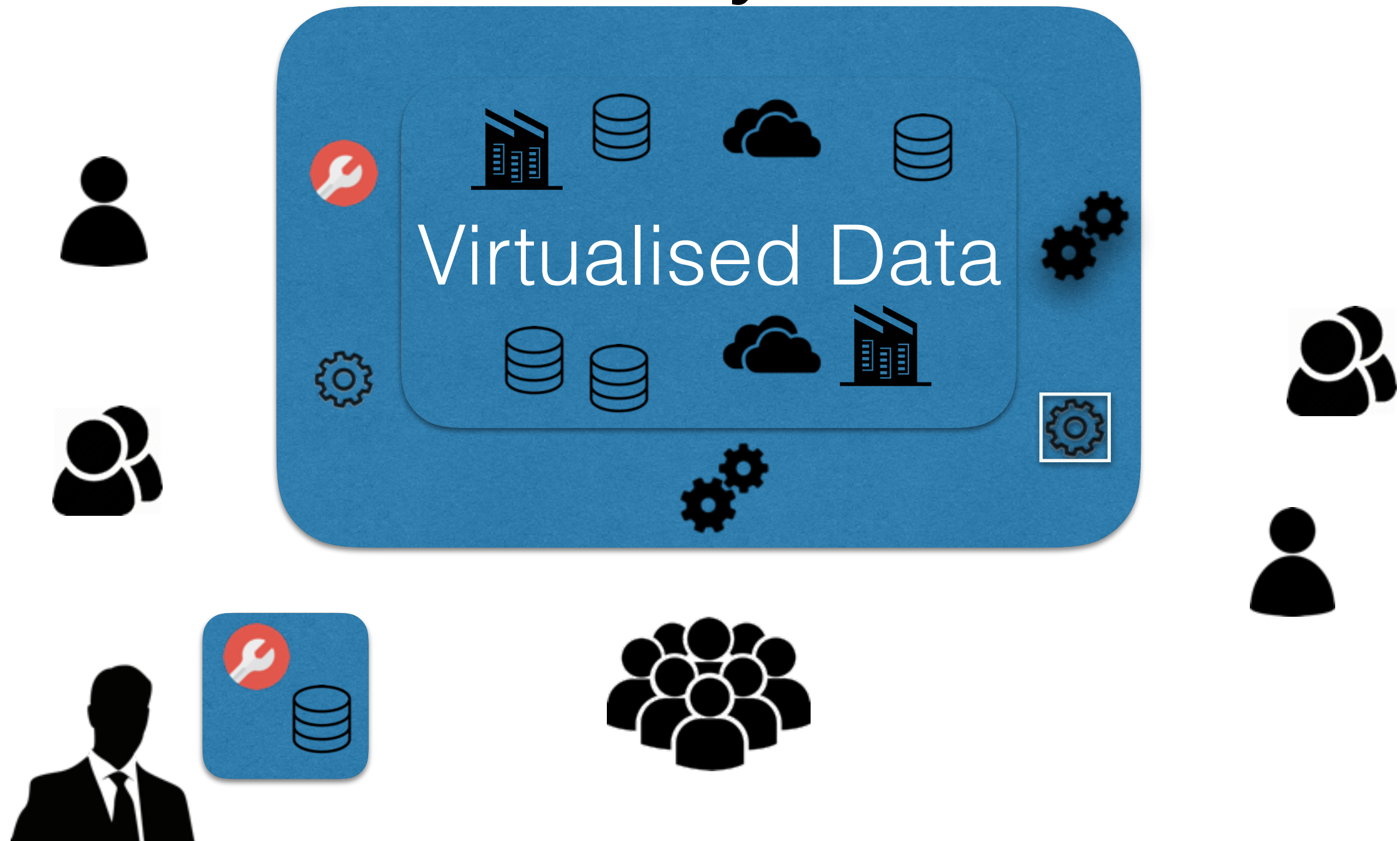
Data Accessibility and Analysis



Data Accessibility and Analysis

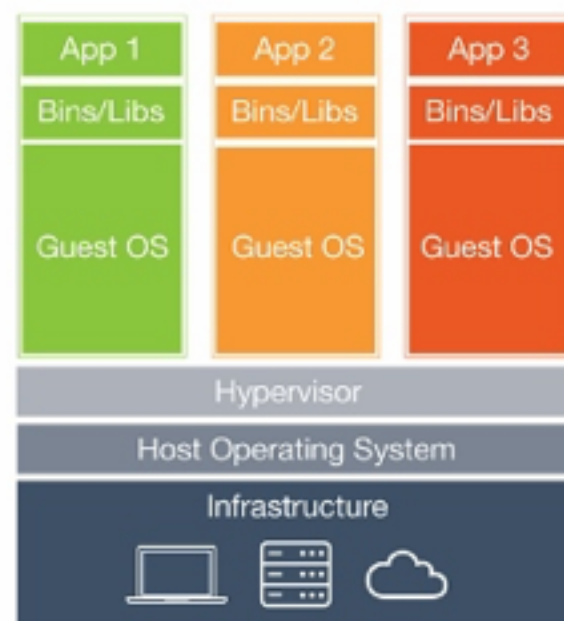


Data Accessibility and Analysis

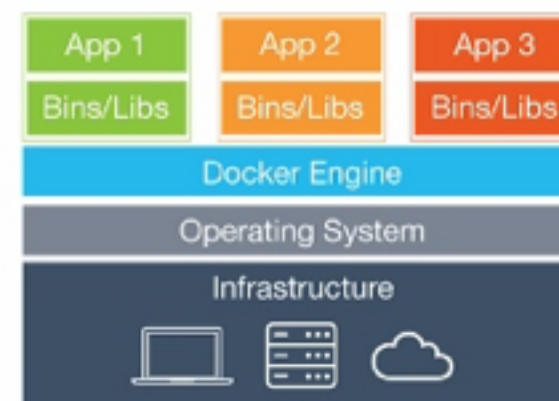




- Docker is an open-source project that automates the deployment of Linux applications inside software containers
- Docker provides an additional layer of abstraction and automation of operating-system-level virtualisation on Linux
- Docker can be integrated into various infrastructure tools, including Amazon Web Services, Google Cloud Platform, OpenStack Nova, etc
- Containers running on a single machine share the same operating system kernel; they start instantly and use less RAM. Images are constructed from layered filesystems and share common files, making disk usage and image downloads much more efficient.



Virtual Machines



Containers

NGS_{easy}



+



=



Install it:

```
:~$ git clone https://github.com/KHP-Informatics/ngseasy.git  
:~$ cd ngseasy  
:~$ make INSTALLDIR="/media/Data" all  
:~$ sudo make install
```

Run it:

```
:~$ cd /media/Data/ngs_projects/congif_files  
:~$ ngseasy -c my_config.tsv -d /media/Data/ngs_projects
```

<https://github.com/KHP-Informatics/ngseasy>

What is on the plate

- An iRODS system tailored for ALS research needs

SECURE
COLLABORATION



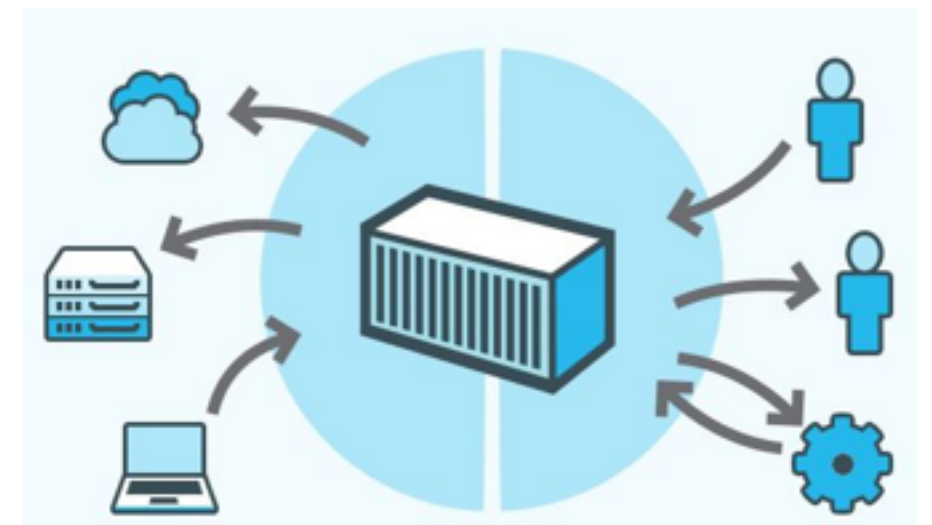
WORKFLOW
AUTOMATION



DATA
DISCOVERY

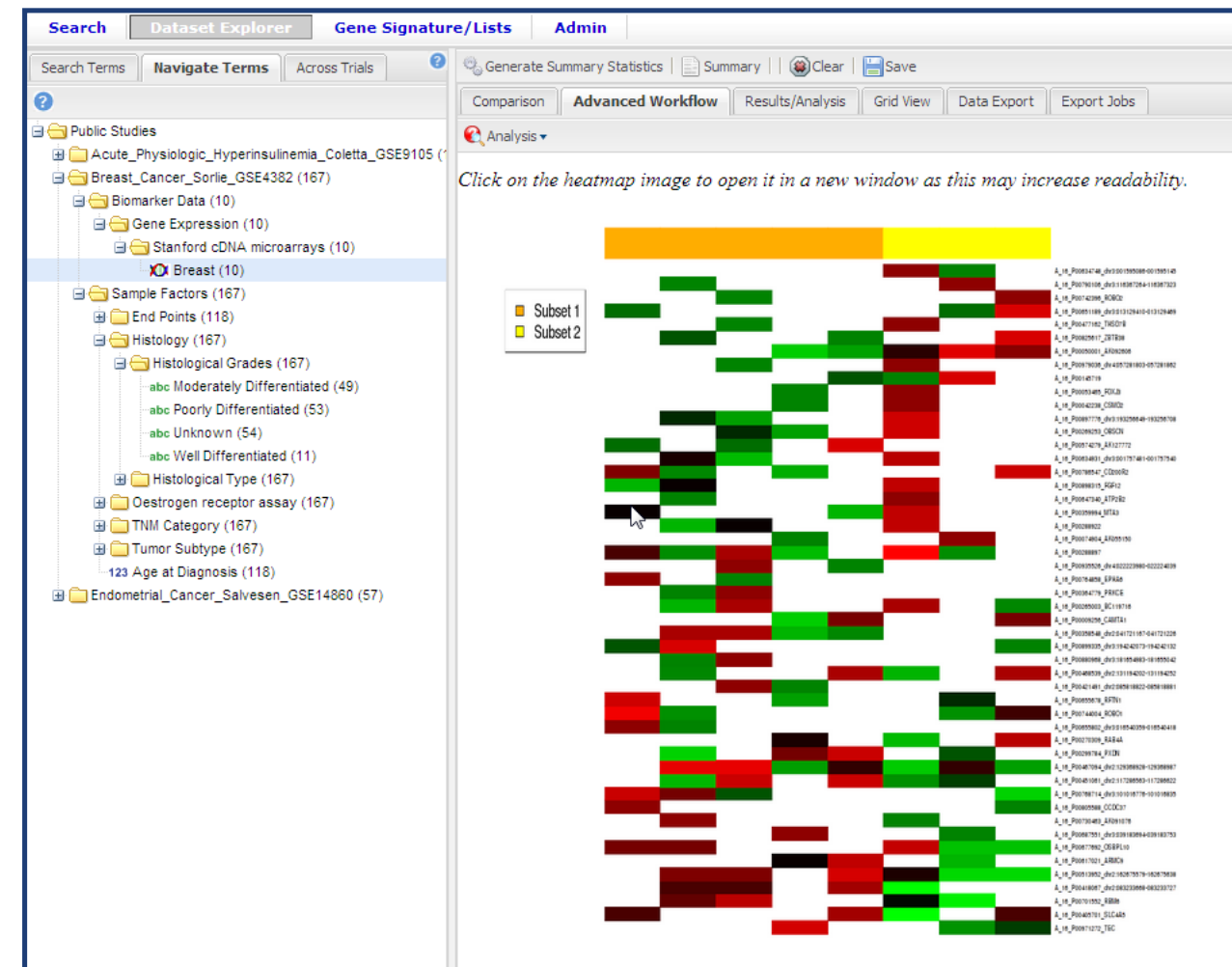


- Dockerized analysis pipelines





- open-source, community-driven knowledge management platform for translational medicine
- organises clinical and research data on per patient base allowing:
 - Compare data from proteomics, metabolomics and other “omics” studies
 - Contrast patterns of gene expression in healthy and diseased individuals and in human tissue samples
 - Investigate correlations between genotype and phenotype in clinical trial data
 - Mine pre-clinical data for insights into the biology of human disease
 - Study genetic and environmental factors involved in human disease
 - Display data visually using a graphical interface
 - Stratify clinical data into molecular subtypes of a specific disease
 - Collaborate across academic, government and corporate research sectors



Ok! All good but who is going to make this happen? and how?

SERVICE	Deployment time	Needed experience
iRODS local	hours	basic
Docker iRODS	minutes	basic
Docker	minutes	basic
Docker container (r)	minutes	basic
Docker image (r)	minutes	basic
iRODS rules (r)	hours	intermediate
iRODS rules (w)	hours	needs training
iRODS administration	minutes	needs training
Docker container (w)	hours	needs training
Docker image (w)	hours	needs training

Summary

A. Data management iRODS

- data sharing
- data accessibility
- data curation
- automatise workflow
- exploit metadata potentialities

B. docker, standardised pipelines, etc

- docker files
- docker images
- shared scrips
- Github

C. Community driven project

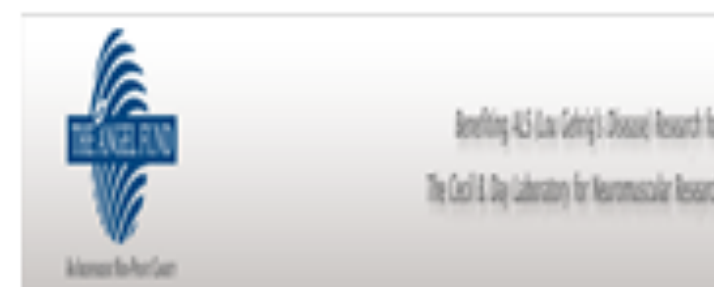
- tailor the system according to our needs

Future Plans

- TranSMART platform interactive analysis
- Develop portable analysis pipelines
- iRODS management system refinement
- User friendly interface



Leading science for better health



Economic and Social Research Council
Shaping Society





Clinical Neuroscience

- Ammar Al-Chalabi
- Ahmad Al Khleifat
- Aleksey Shatunov
- Anna Kulka
- Anand Pandit
- Ashley Jones
- Sarah Martin
- William Sproviero

Health Informatics

- Stephen J Newhouse
- Richard Dobson

All our collaborators!!!!!!!!!!

