eTRIKS Project Sneak Peek

FJPPL Computing Workshop @ CCIN2P3, March 11th 2015

Benjamin Guillon, Systems Engineer

CCIN2P3

eTRIKS

European Translational Information and Knowledge Management Services

efpia

imi

cnrs

Agenda

Project Overview

- Translational research
- The eTRIKS project

Platform Overview

- A cloud based platform
- Project Modules
- Key Services
- Hosted Projects

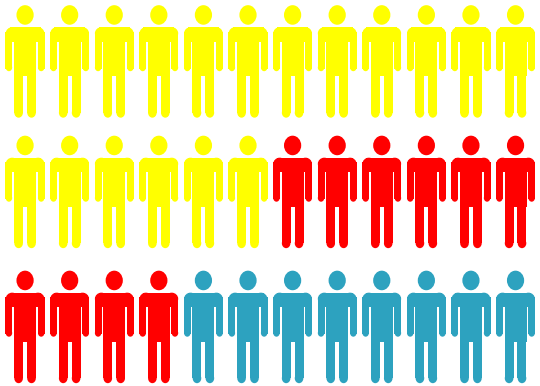
Platform Usage

- Platform Monitoring
- Usage Statistics

Next Steps

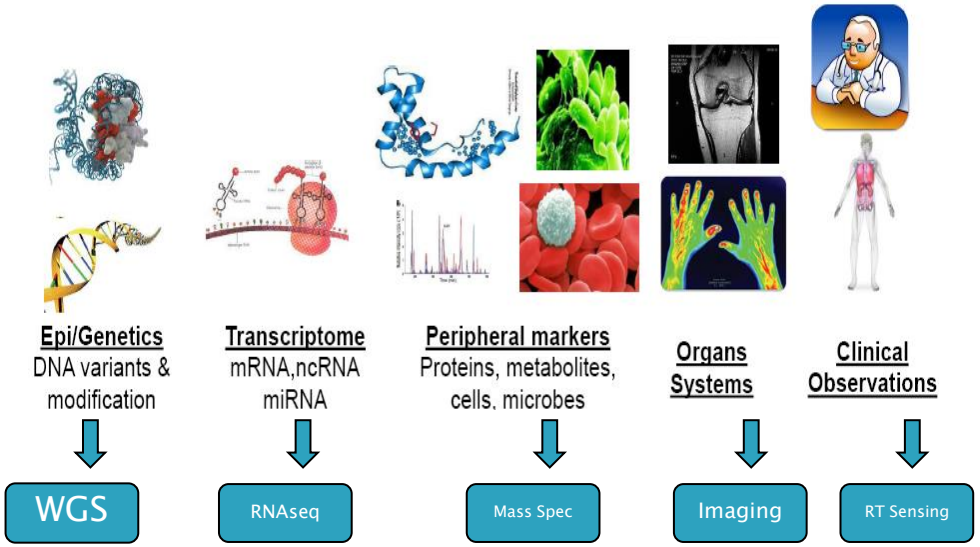
Translational Research

Patients with diseases



+

Bioassays: measurements on genes, molecules, organs



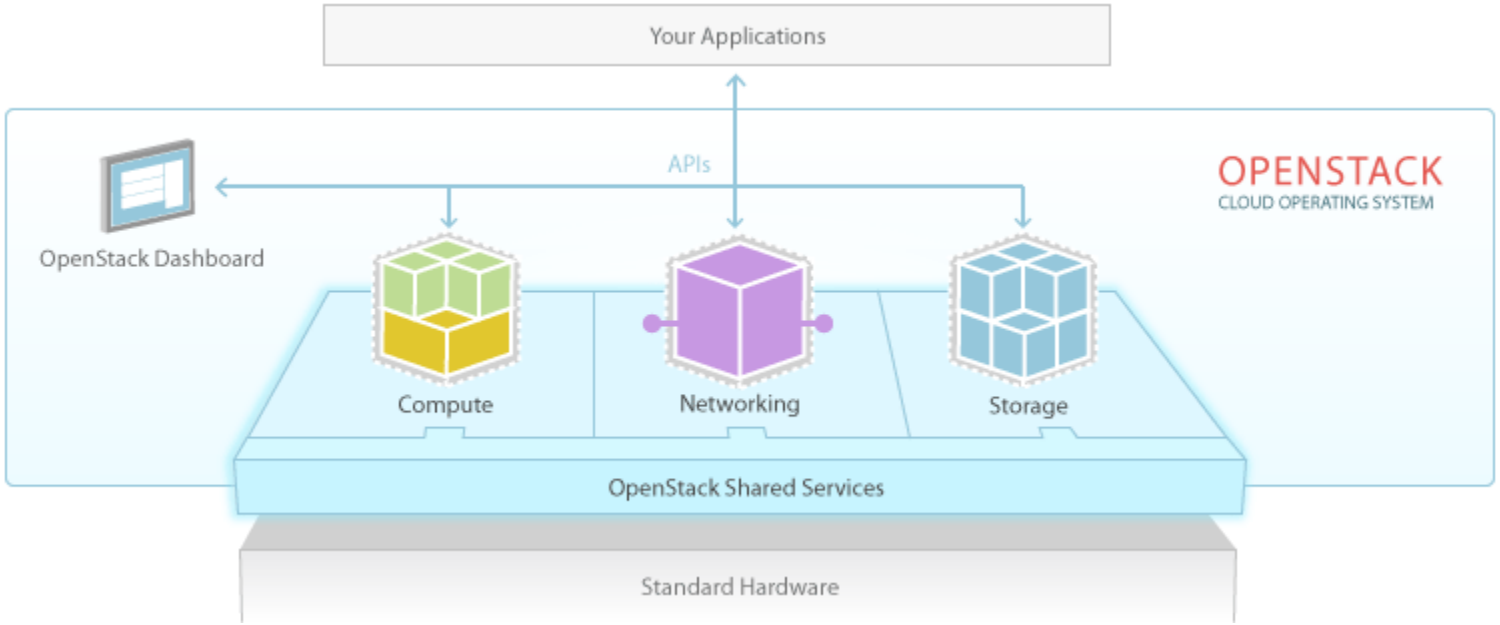
- Combine clinical observations and bioassay techniques
  - provide more efficient research of treatments

- A knowledge management platform
  - Store, curate, analyze and export data
  - Host multiple IMI projects (and others…)

- Increasing the efficiency of translational research
  - Reduced costs: one platform to rule them all
  - Cross study analyses
  - Private/Public Translational Research IMI support
  - Open agreed standards across the IMI TR projects
  - Open source and interoperable software: TranSMART

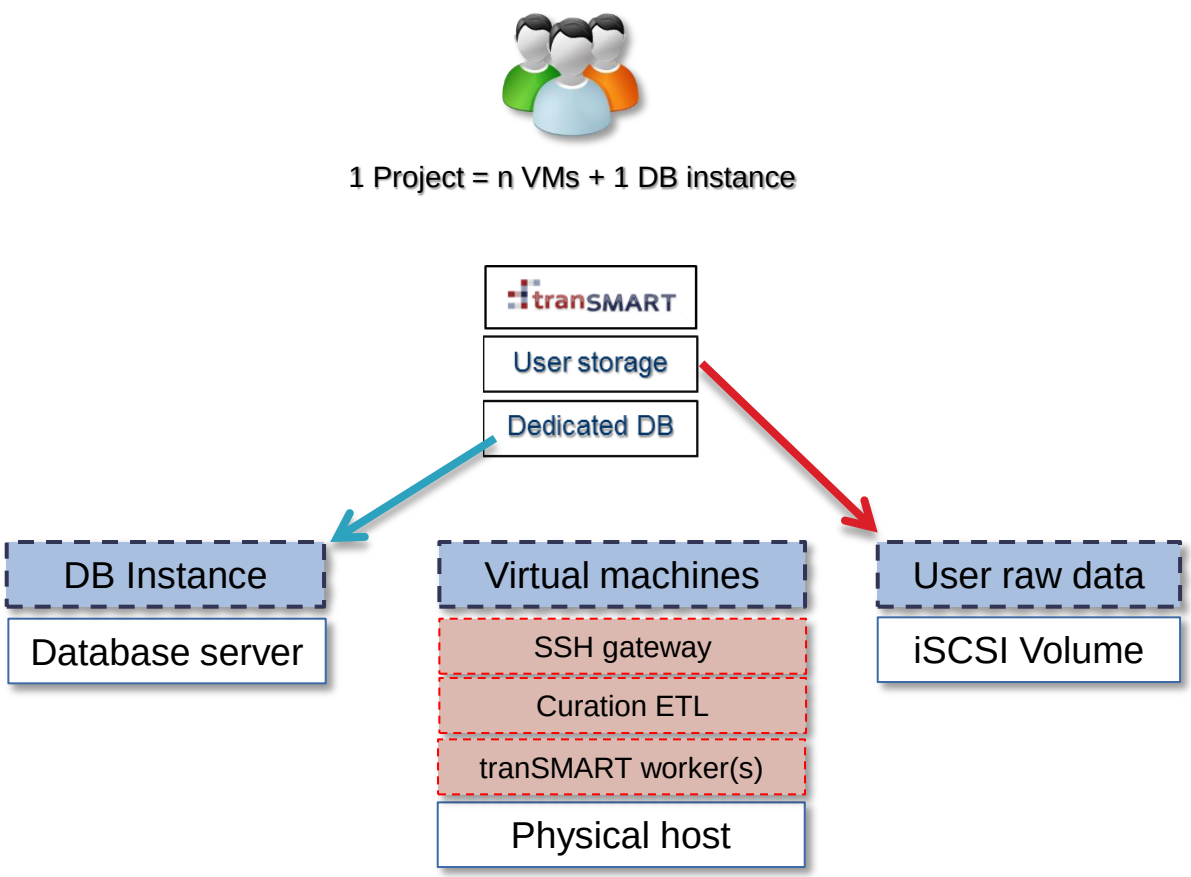
Innovative Medicines Initiative

A cloud based platform

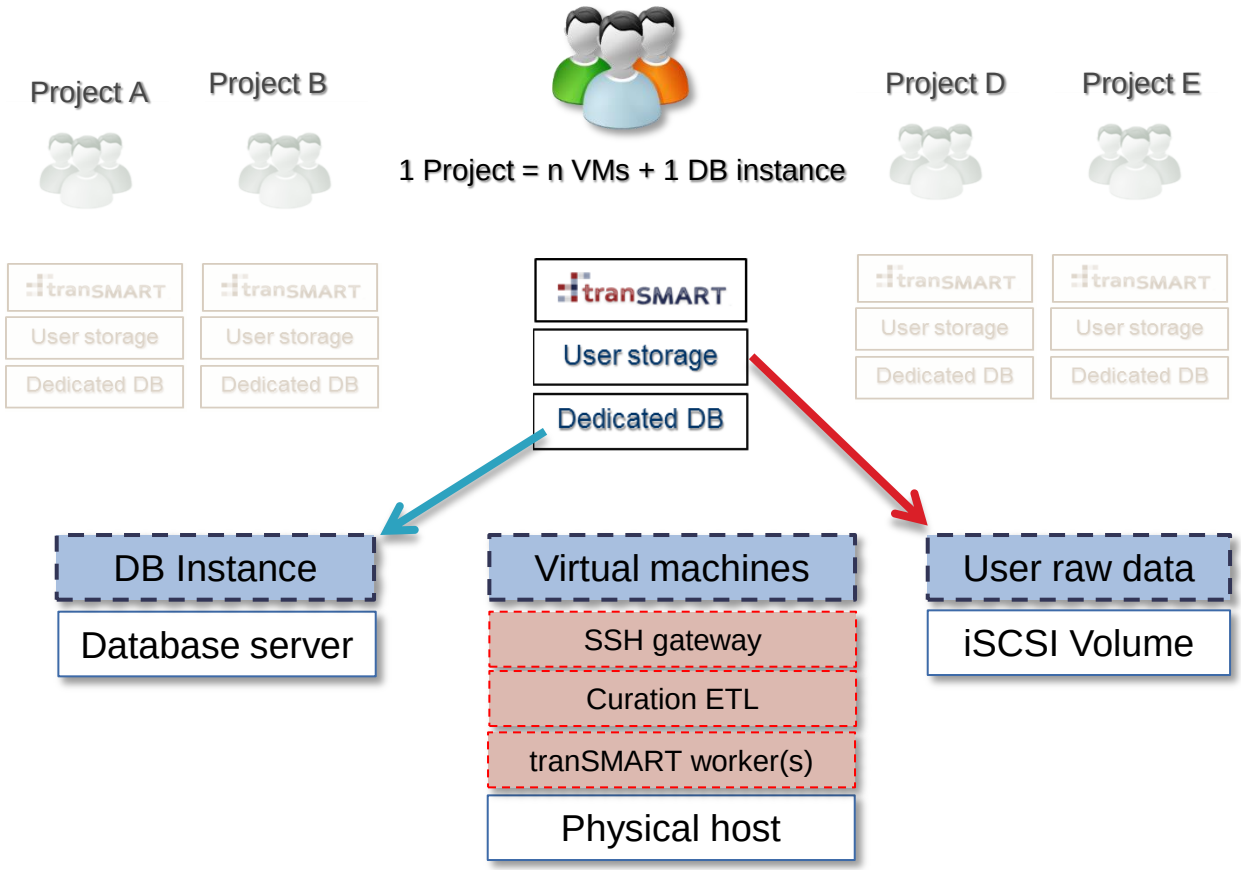


- Quick provisioning
- Horizontal Scalability
- Resources utilization efficiency

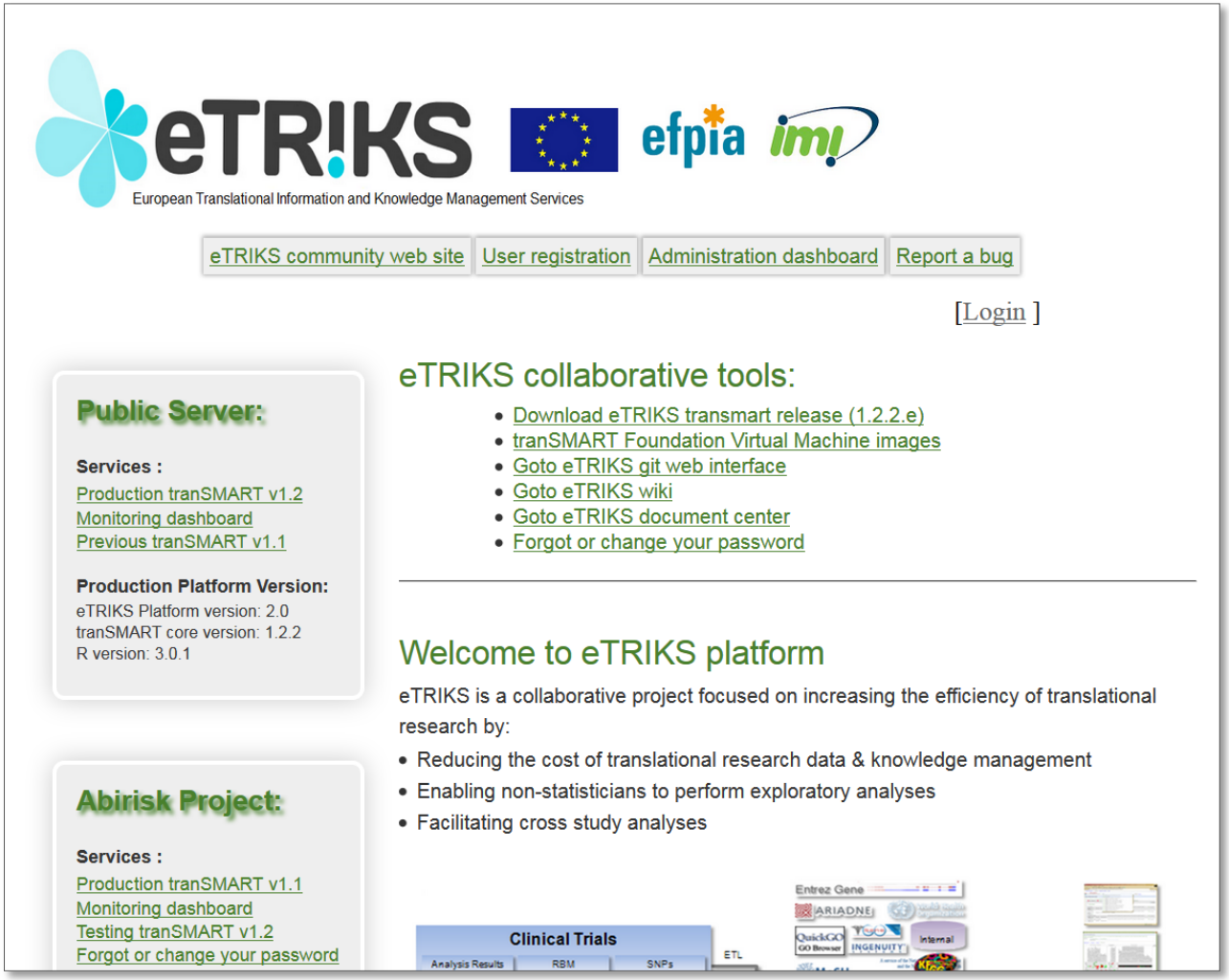
Project Modules (1)



Project Modules (2)

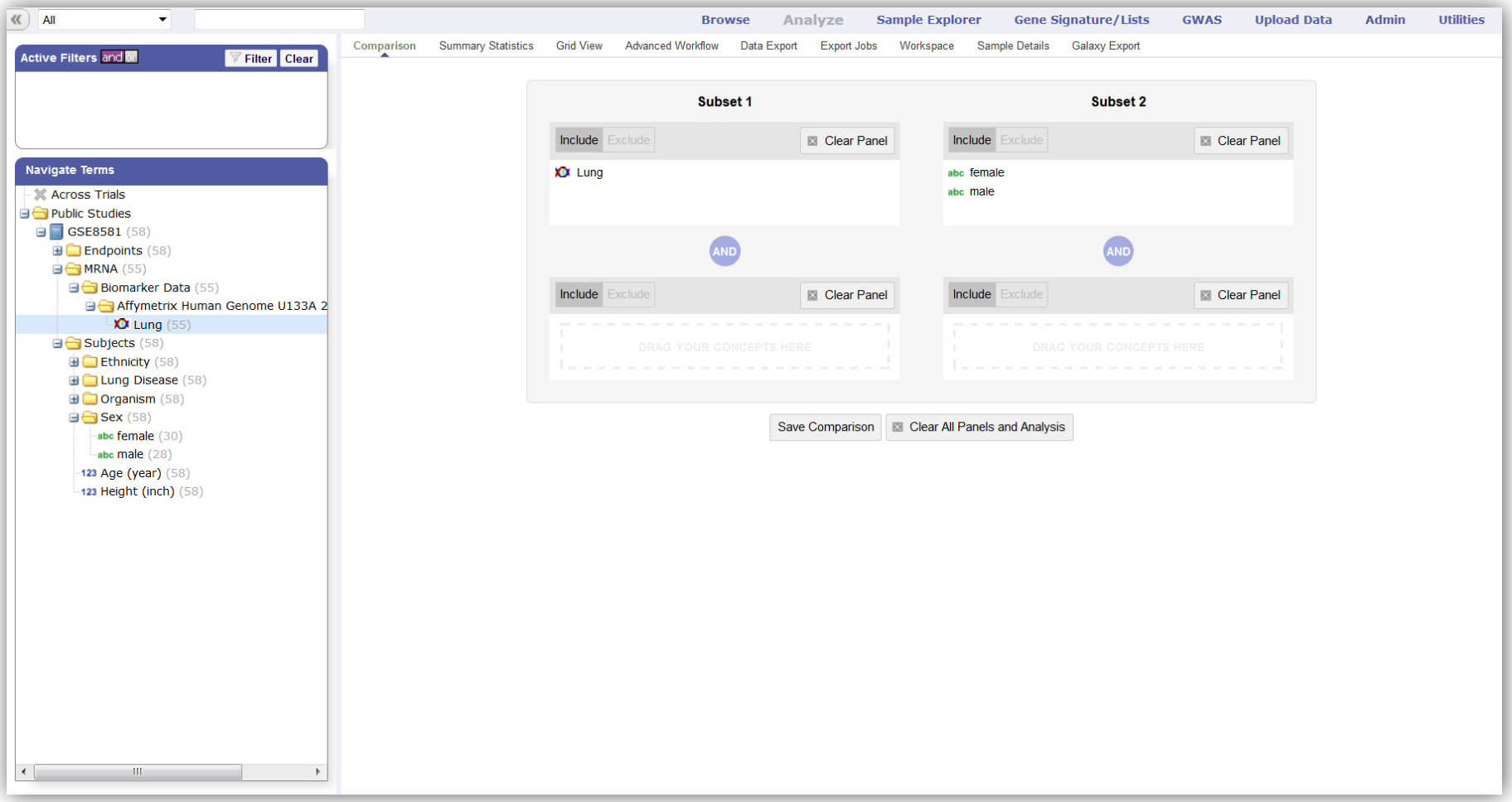


- The eTRIKS web portal

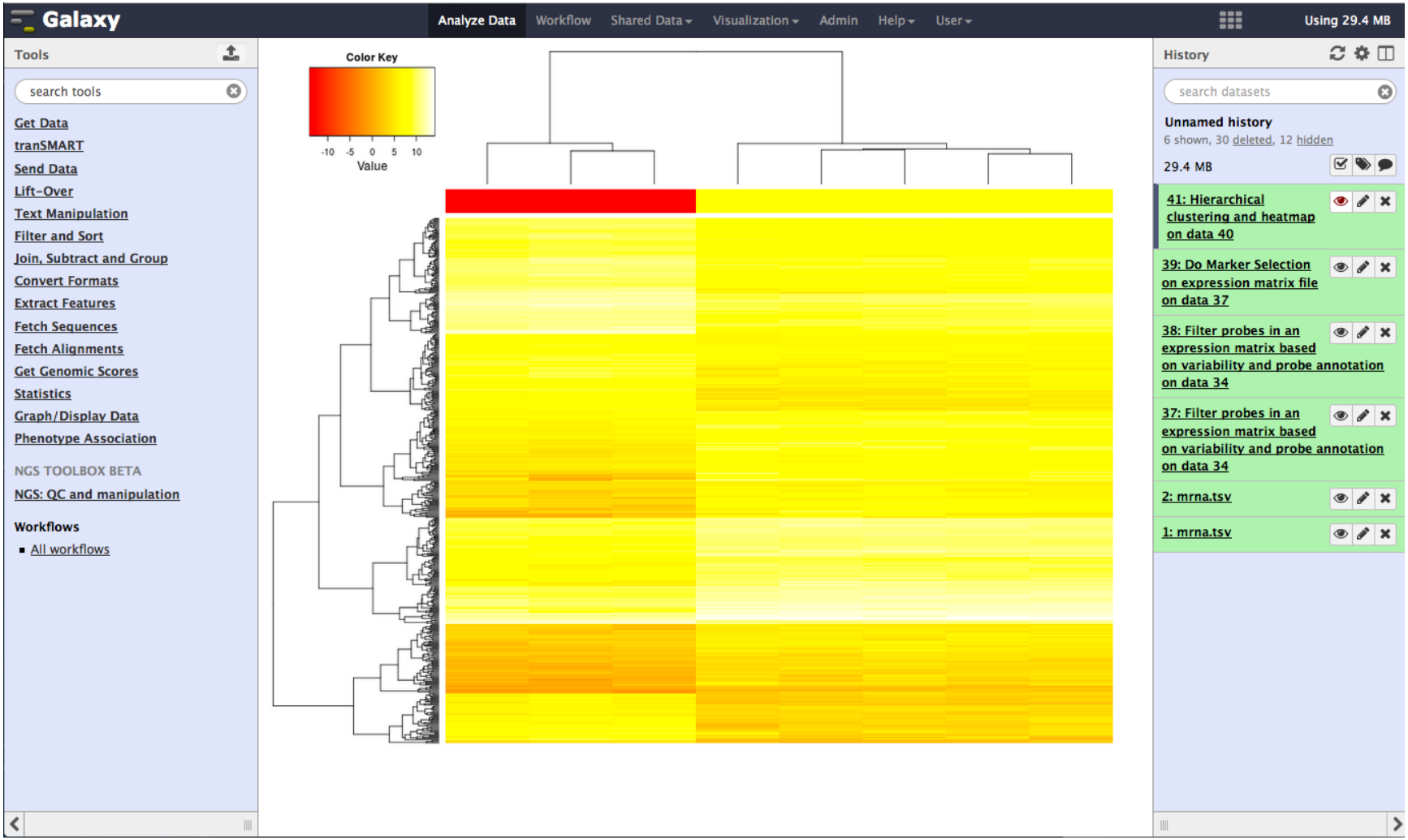


Key Services (2)

- TranSMART



- Galaxy


Galaxy
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
Using 29.4 MB
Tools
search tools
Get Data
tranSMART
Send Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Statistics
Graph/Display Data
Phenotype Association
NGS TOOLBOX BETA
NGS: QC and manipulation
Workflows
All workflows
Color Key
-10 -5 0 5 10
Value
History
search datasets
Unnamed history
6 shown, 30 deleted, 12 hidden
29.4 MB
41: Hierarchical clustering and heatmap on data 40
39: Do Marker Selection on expression matrix file on data 37
38: Filter probes in an expression matrix based on variability and probe annotation on data 34
37: Filter probes in an expression matrix based on variability and probe annotation on data 34
2: mrna.tsv
1: mrna.tsv

Key Services (4)

- Data Curation Environment
  - Dedicated curation VM
  - ETL tools
  - Direct database access
  - Complete Linux environment for curators

- Data Storage Services
  - Various types (Block storage, Databases)
  - 240TB of raw storage space
  - Focused on reliability

Hosted Projects

- Oncotrack
  - Oncology research
  - www.oncotrack.eu

- Abirisk
  - Drug immunization research
  - www.abirisk.eu

- Public Server: the eTRIKS showroom
  - Accessible from the portal
  - Open access
  - Public data from clinical studies on various diseases

- More projects joining soon…

Multiple levels of monitoring

End-users

- Application access metrics
- Various information about users
- Time spent using the applications

AWStats

Services

- Apache web servers
- Tomcat application servers and the Java Virtual Machine
- PostgreSQL databases
- SSH gateways

MUNIN

kibana

logstash

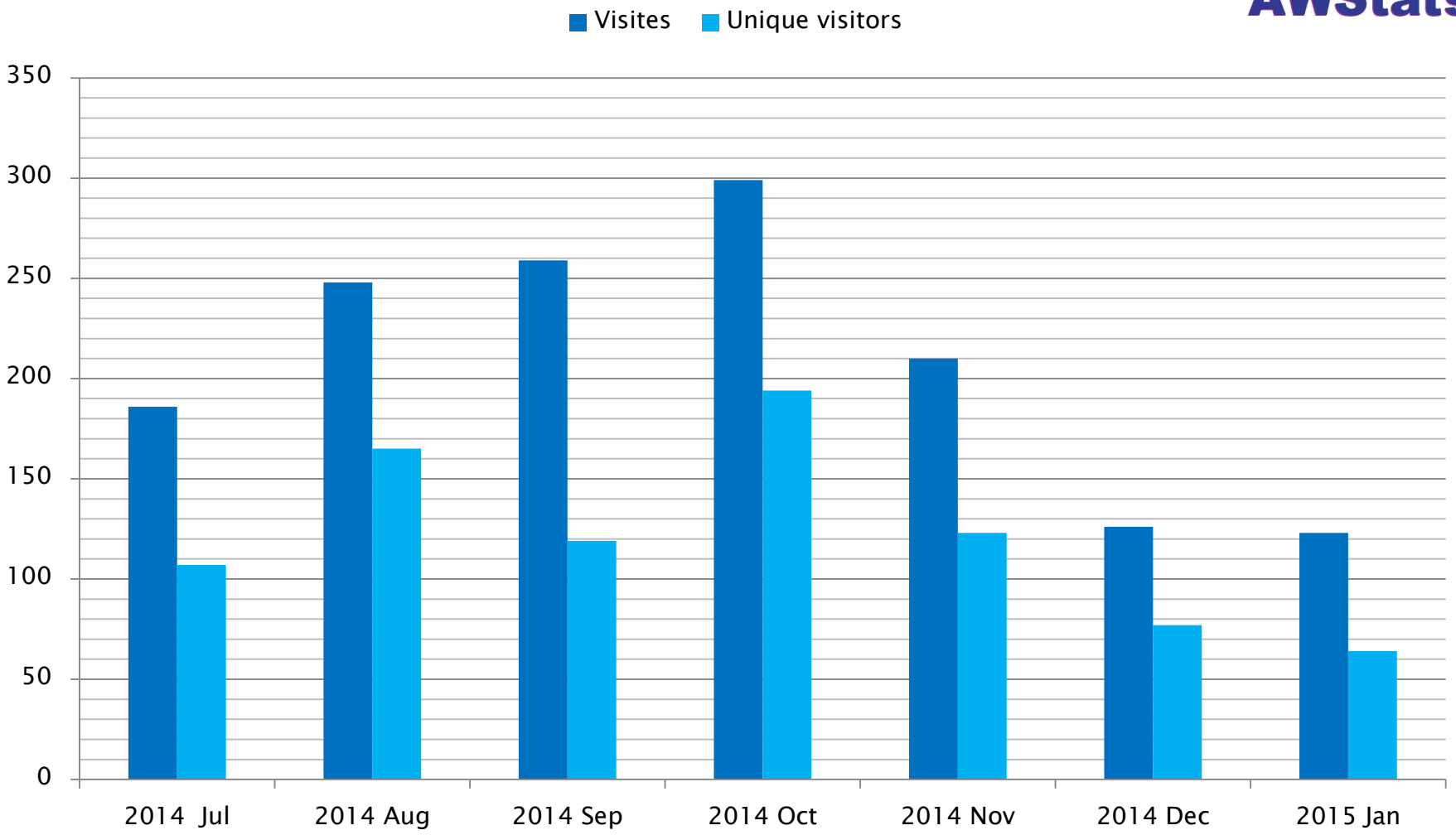
Infrastructure

- Compute and storage metrics
- CPU, disk I/O, memory and network
- Storage space capacity
- Electrical consumptions

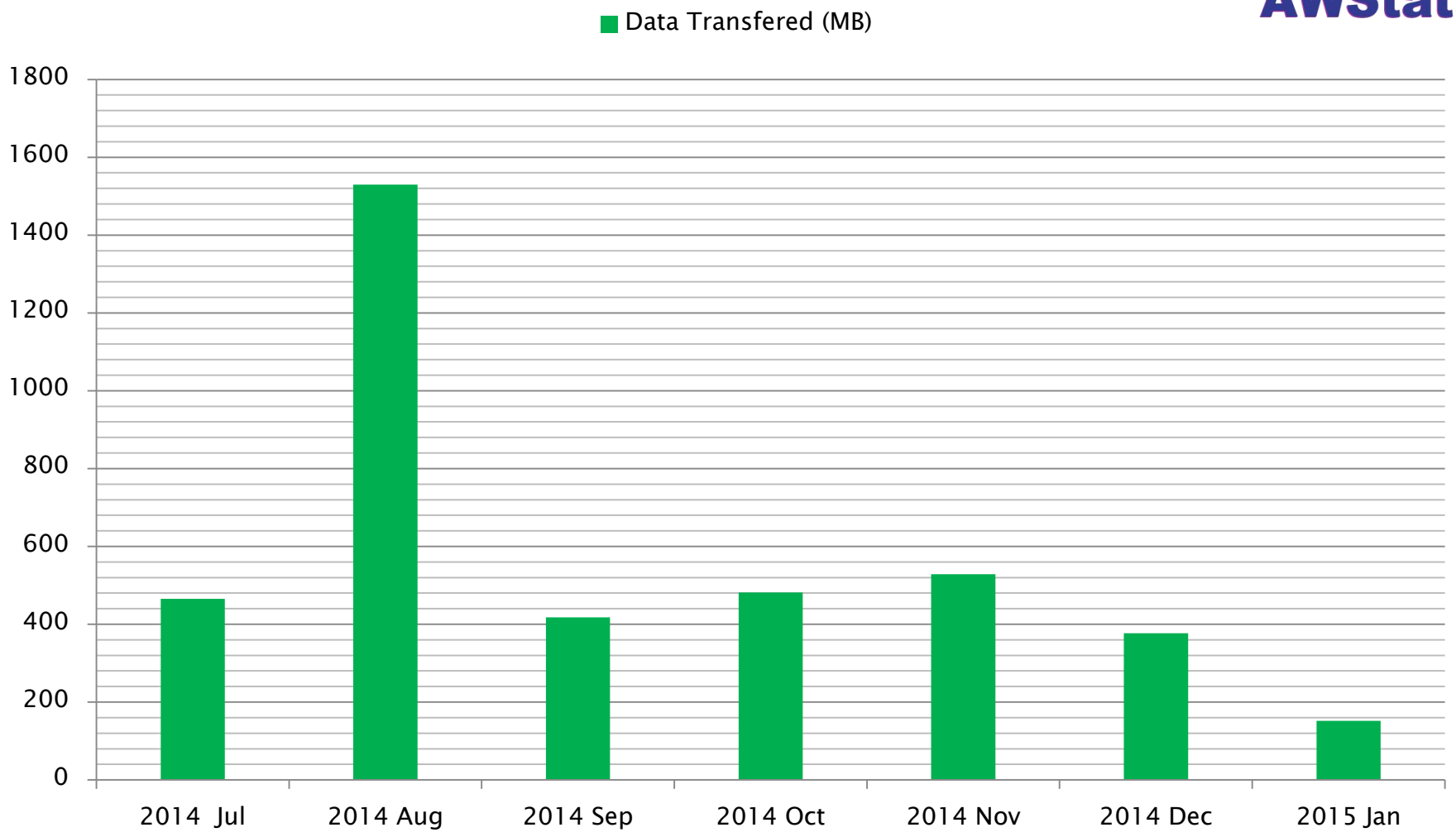
collectd
The system statistics collection daemon

Nagios®

- Users on the public server

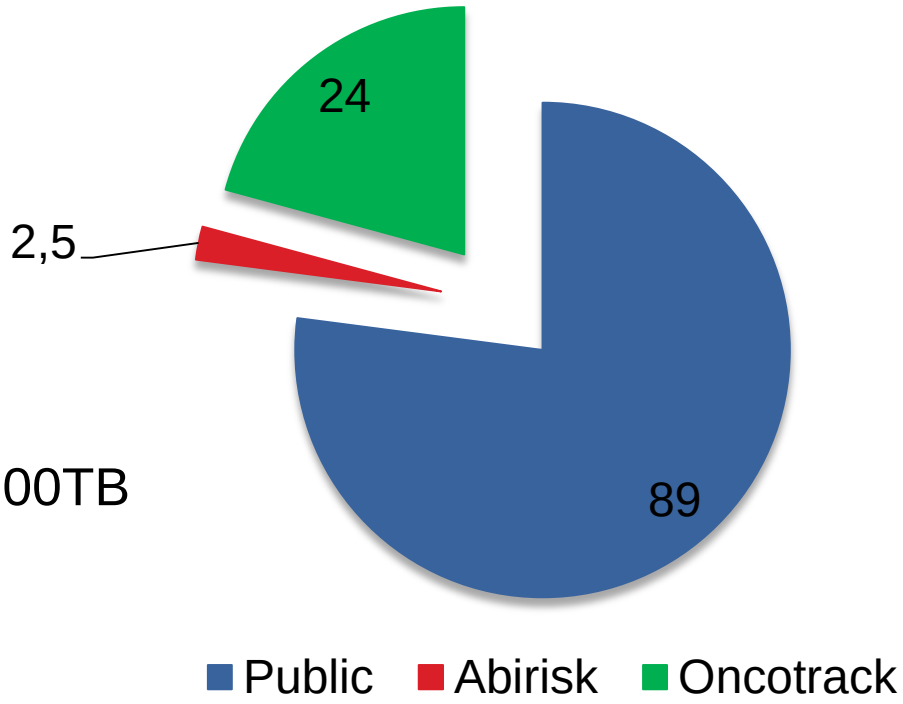


- Users on the public server



- Database cumulated sizes

  Addition of the TranSMART v1.1, v1.2 (and Galaxy when available) databases.



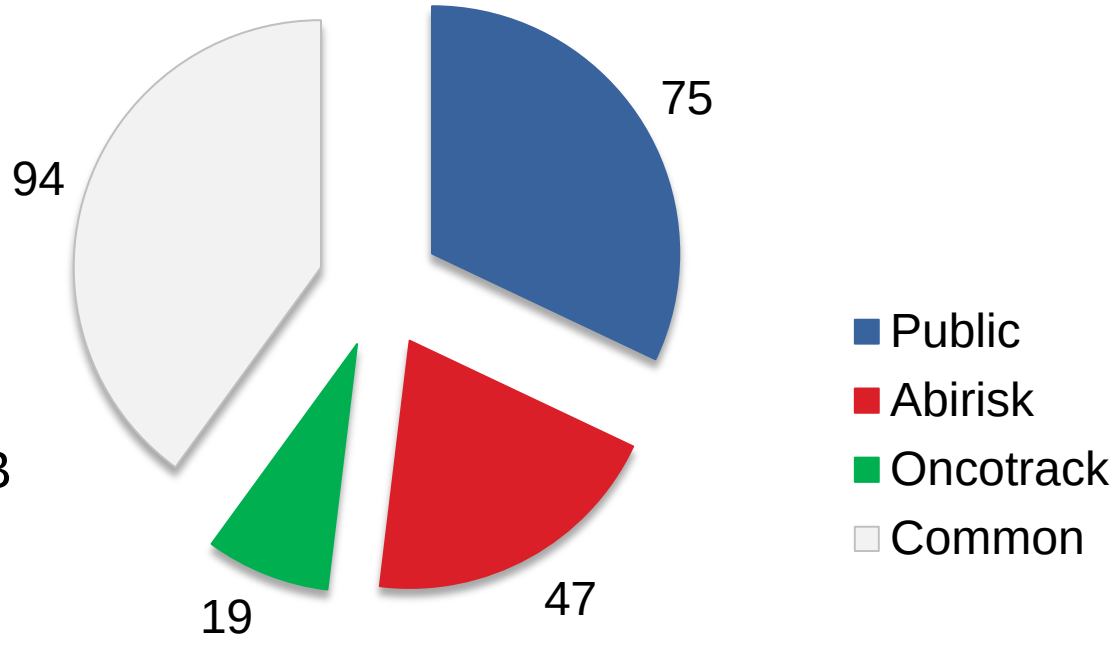
  - ~ 115GB / 100TB

PostgreSQL

Disk space usage

Addition of all the projects disk storage used (excluding databases).

- ~ 235GB / 100TB



Usage in gigabytes (GB)

Next Steps

- Multi-site federated cloud
  - Horizontal scaling
  - Fail-over and high availability
  - Data backup
  - Legal issues
- Application breakdown
  - From a monolithic to a modular architecture
  - Increase horizontal scalability for the backends
    - PostgreSQL, R, Tomcat.
  - Mutualize frontends
    - Apache, HAProxy.
- Docker support
  - On top of Openstack?
  - For dissemination purposes
- Automate services deployment
  - Use puppet to deploy and maintain services

Any questions?

Next week in eTRIKS …

- Dr. Pengfei Liu
  - eTRIKS software security developer from CCIN2P3
  - ISGC 2015 Taiwan (Taipei) - March 15th - 20th 2015
    - http://event.twgrid.org/isgc2015/