

Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules

eTRIKS Project Sneak Peek

FJPPL Computing Workshop @ CCIN2P3, March 11th 2015

Benjamin Guillon, Systems Engineer



Project Overview

- ▶ Translational research
- ▶ The eTRIKS project

Platform Overview

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

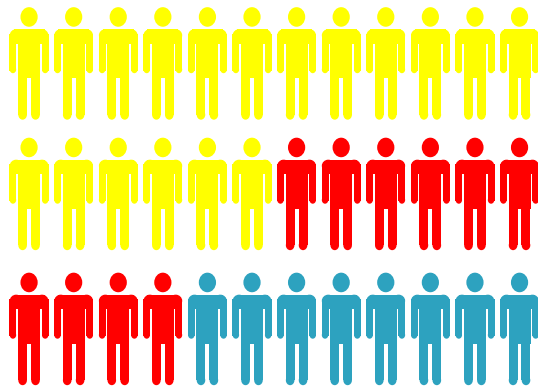
Platform Usage

- ▶ Platform Monitoring
- ▶ Usage Statistics

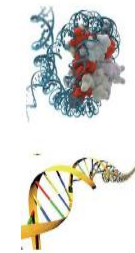
Next Steps

Bioassays: measurements on genes, molecules, organs

Patients with diseases



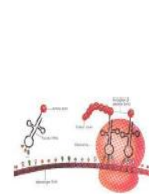
+



EpiGenetics
DNA variants &
modification



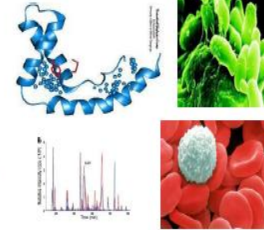
WGS



Transcriptome
mRNA, ncRNA
miRNA



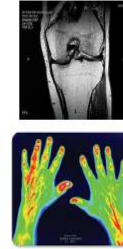
RNAseq



Peripheral markers
Proteins, metabolites,
cells, microbes



Mass Spec



Organs
Systems



Imaging



Clinical
Observations



RT Sensing

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



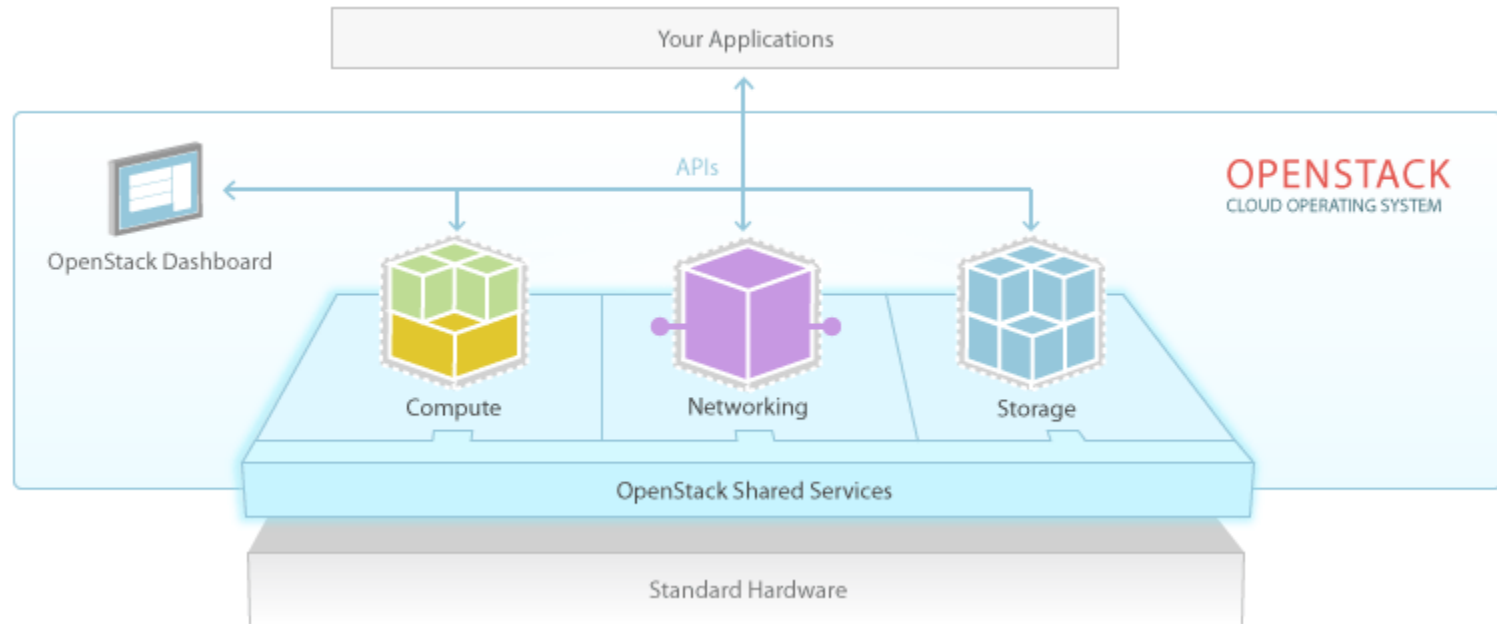
- ▶ A knowledge management platform
 - Store, curate, analyze and export data
 - Host multiple IMI projects (and others...)

- ▶ 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



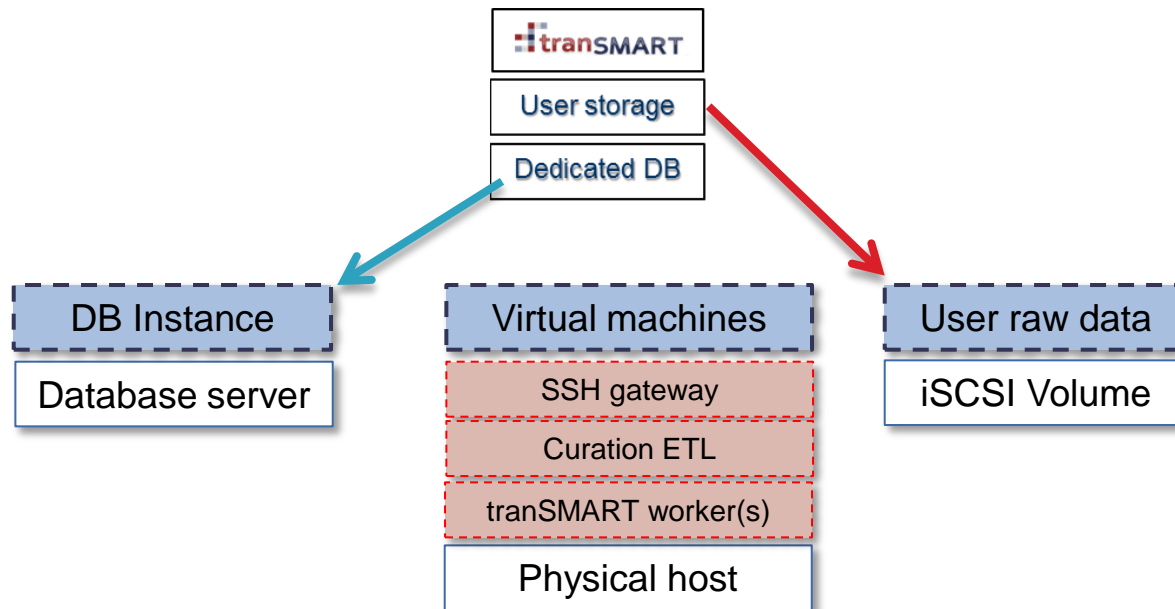


- ▶ Quick provisioning
- ▶ Horizontal Scalability
- ▶ Resources utilization efficiency

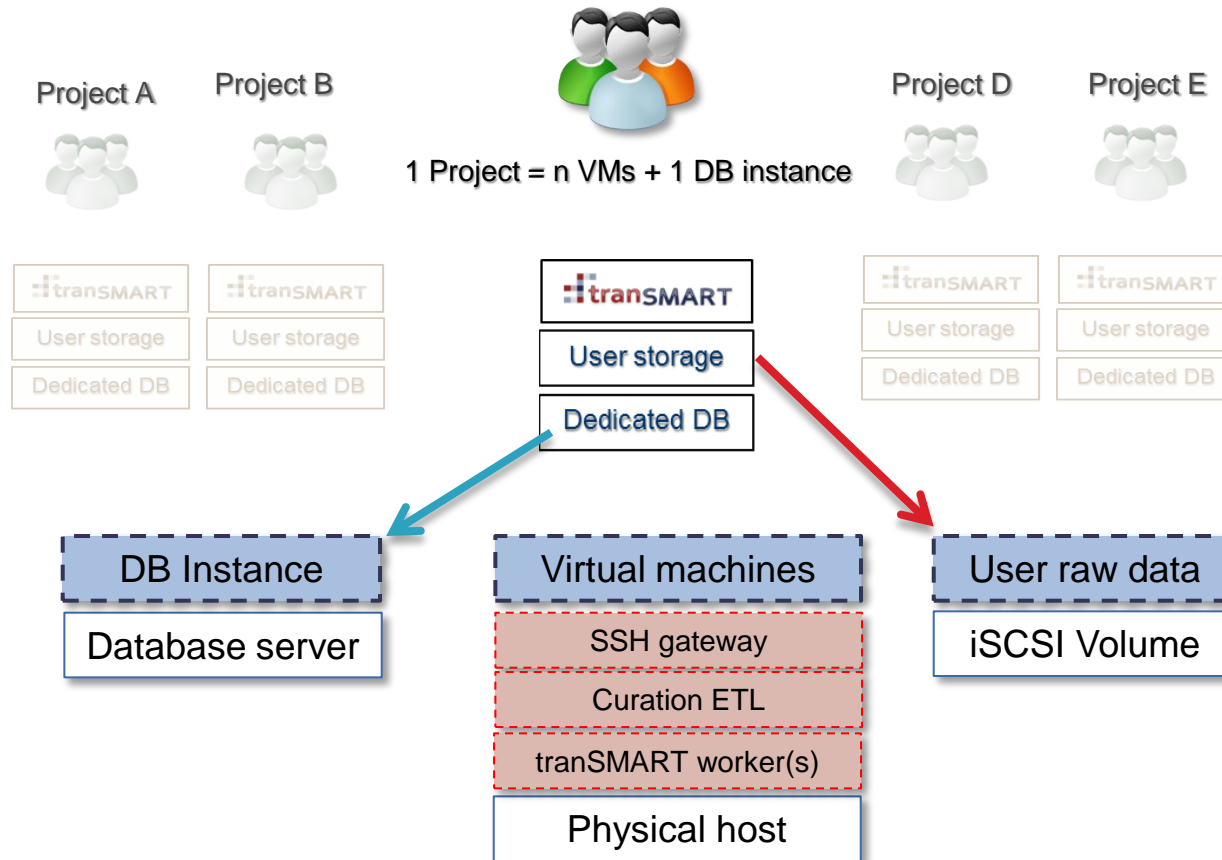
Project Modules (1)



1 Project = n VMs + 1 DB instance



Project Modules (2)



▶ The eTRIKS web portal

The screenshot shows the eTRIKS web portal homepage. At the top left is the eTRIKS logo, a stylized blue flower, followed by the text "eTRIKS" and "European Translational Information and Knowledge Management Services". To the right are logos for the European Union, efpia, and imi. Below the logos is a navigation bar with buttons for "eTRIKS community web site", "User registration", "Administration dashboard", and "Report a bug". A "[Login]" link is positioned to the right of the navigation bar. The main content area is divided into several sections. On the left, there are two grey boxes: "Public Server:" and "Abirisk Project:". The "Public Server:" box lists services like "Production tranSMART v1.2", "Monitoring dashboard", and "Previous tranSMART v1.1", and also provides "Production Platform Version" information. The "Abirisk Project:" box lists services like "Production tranSMART v1.1", "Monitoring dashboard", "Testing tranSMART v1.2", and "Forgot or change your password". The central section is titled "eTRIKS collaborative tools:" and lists links for downloading releases, virtual machine images, accessing the git interface, wiki, document center, and password management. Below this is a "Welcome to eTRIKS platform" section with a brief description and a list of project goals: reducing research costs, enabling non-statisticians, and facilitating cross-study analyses. At the bottom, there are several small icons for various tools and services, including Entrez Gene, ARIADNE, QuickGO, GO Browser, INGENUITY, and Internal.

eTRIKS
European Translational Information and Knowledge Management Services

[eTRIKS community web site](#) [User registration](#) [Administration dashboard](#) [Report a bug](#)

[\[Login\]](#)

Public Server:

Services :
[Production tranSMART v1.2](#)
[Monitoring dashboard](#)
[Previous tranSMART v1.1](#)

Production Platform Version:
eTRIKS Platform version: 2.0
tranSMART core version: 1.2.2
R version: 3.0.1

Abirisk Project:

Services :
[Production tranSMART v1.1](#)
[Monitoring dashboard](#)
[Testing tranSMART v1.2](#)
[Forgot or change your password](#)

eTRIKS collaborative tools:

- [Download eTRIKS transmart release \(1.2.2.e\)](#)
- [tranSMART Foundation Virtual Machine images](#)
- [Goto eTRIKS git web interface](#)
- [Goto eTRIKS wiki](#)
- [Goto eTRIKS document center](#)
- [Forgot or change your password](#)

Welcome to eTRIKS platform

eTRIKS is a collaborative project focused on increasing the efficiency of translational research by:

- Reducing the cost of translational research data & knowledge management
- Enabling non-statisticians to perform exploratory analyses
- Facilitating cross study analyses

Clinical Trials
Analysis Results | RBM | SNPs | ETL

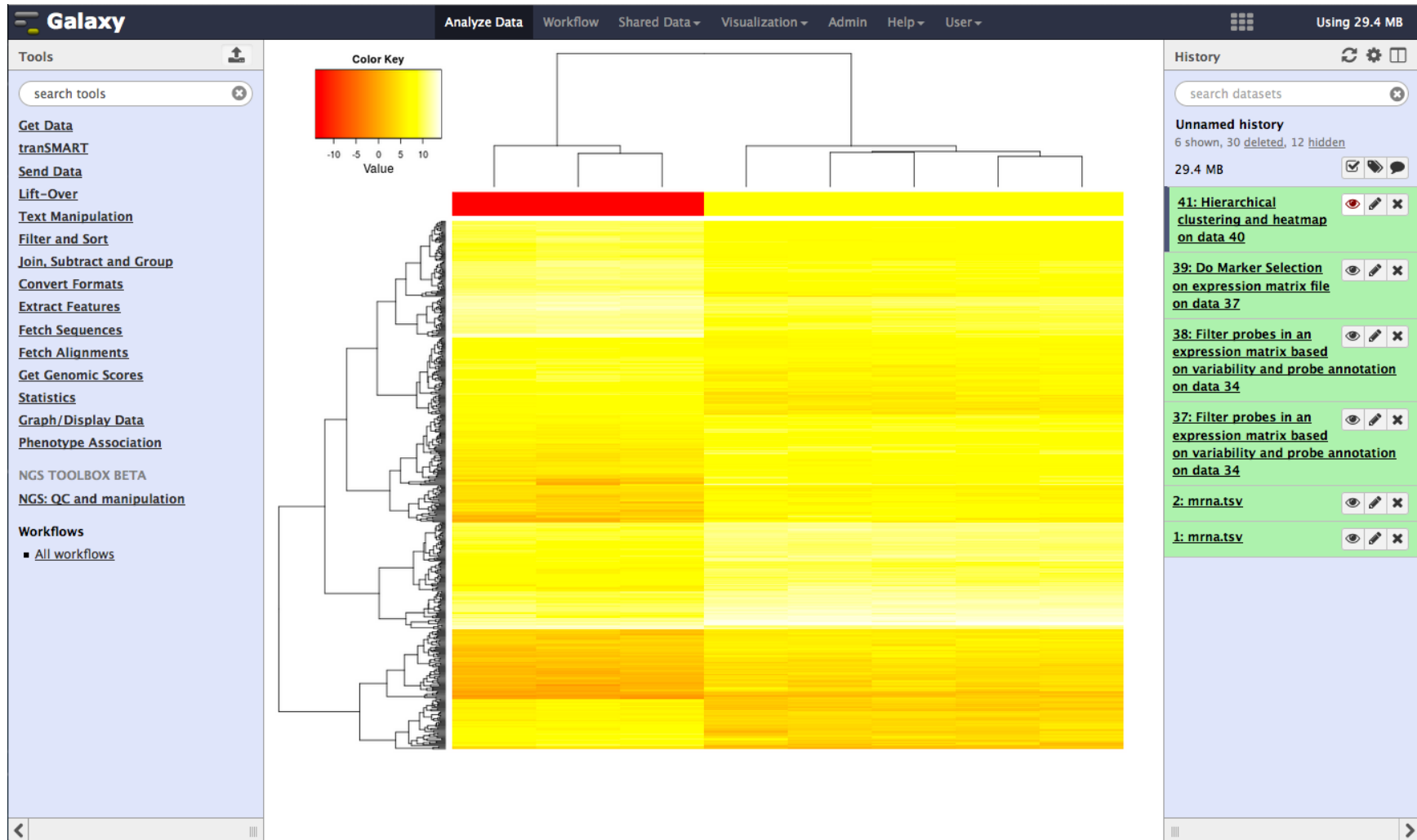
Entrez Gene | ARIADNE | QuickGO | GO Browser | INGENUITY | Internal

▶ TranSMART



The screenshot displays the TranSMART web interface. At the top, there is a navigation bar with tabs: "Browse", "Analyze", "Sample Explorer", "Gene Signature/Lists", "GWAS", "Upload Data", "Admin", and "Utilities". Below this is a secondary navigation bar with options: "Comparison", "Summary Statistics", "Grid View", "Advanced Workflow", "Data Export", "Export Jobs", "Workspace", "Sample Details", and "Galaxy Export". On the left side, there is a "Navigate Terms" panel with a tree view. The tree is expanded to show "Subjects (58)" > "Sex (58)" > "abc female (30)" and "abc male (28)". The "Lung (55)" term is also visible under "Public Studies". The main area shows a comparison setup with two columns, "Subset 1" and "Subset 2". Each column has two panels. The top panel in Subset 1 contains the term "Lung". The top panel in Subset 2 contains the terms "abc female" and "abc male". A blue "AND" button is positioned between the two columns. Below the panels are two dashed boxes with the text "DRAG YOUR CONCEPTS HERE". At the bottom of the comparison area are two buttons: "Save Comparison" and "Clear All Panels and Analysis".

▶ Galaxy



▶ 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



▶ 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



Services

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

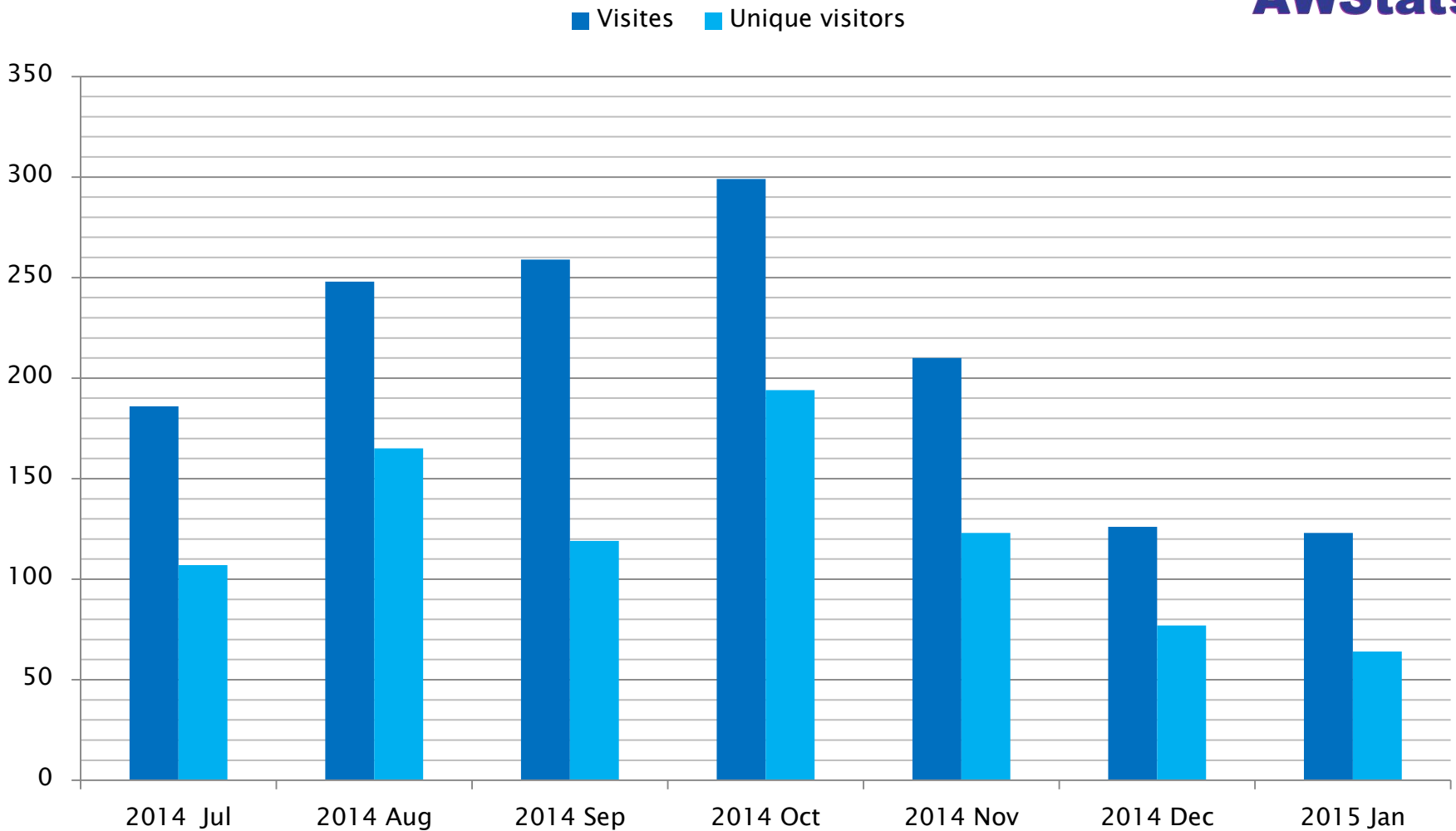


Infrastructure

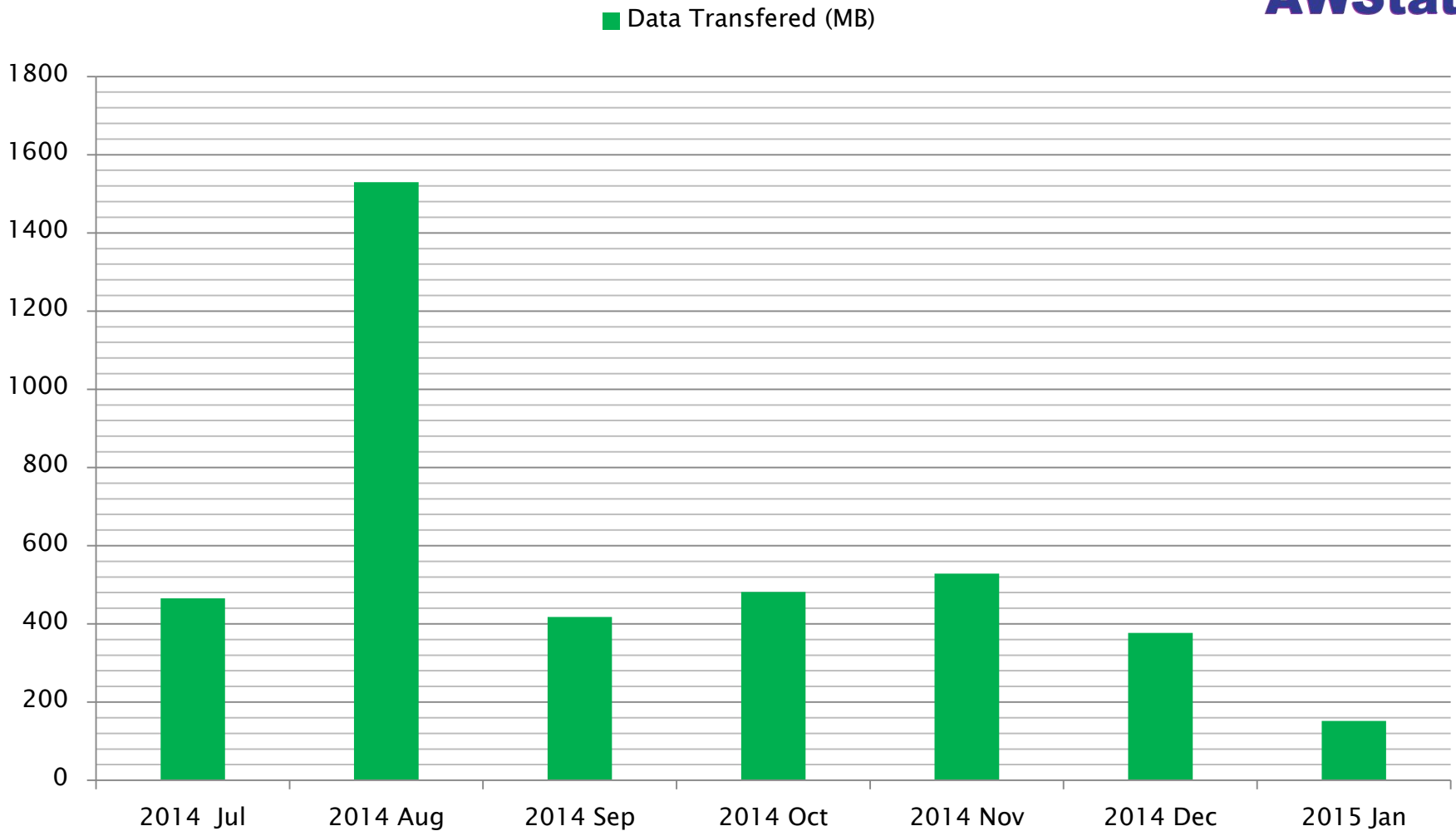
- Compute and storage metrics
- CPU, disk I/O, memory and network
- Storage space capacity
- Electrical consumptions



► Users on the public server

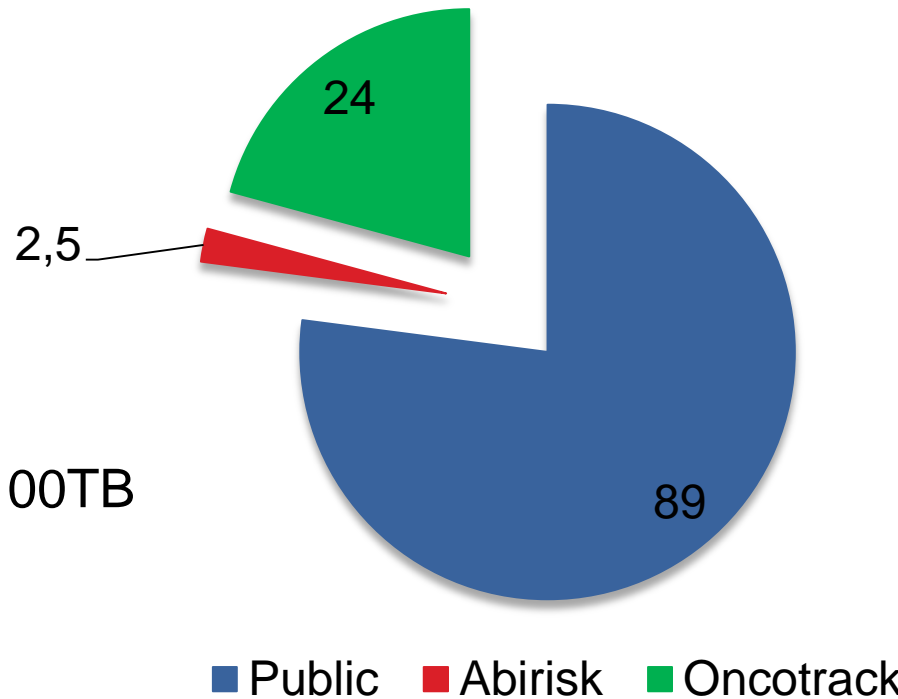


▶ Users on the public server



▶ Database cumulated sizes

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



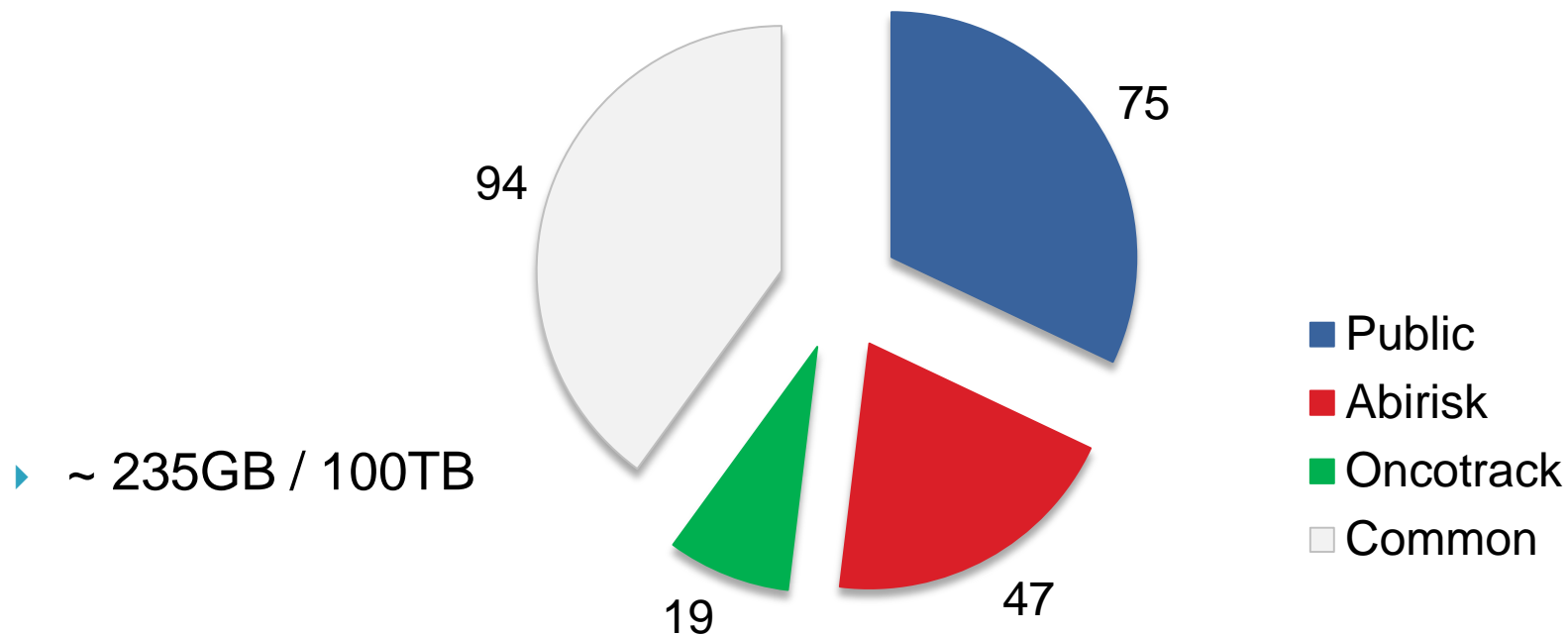
▶ ~ 115GB / 100TB

Loads in gigabytes (GB)



▶ Disk space usage

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



▶ ~ 235GB / 100TB

Usage in gigabytes (GB)



- ▶ **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/>