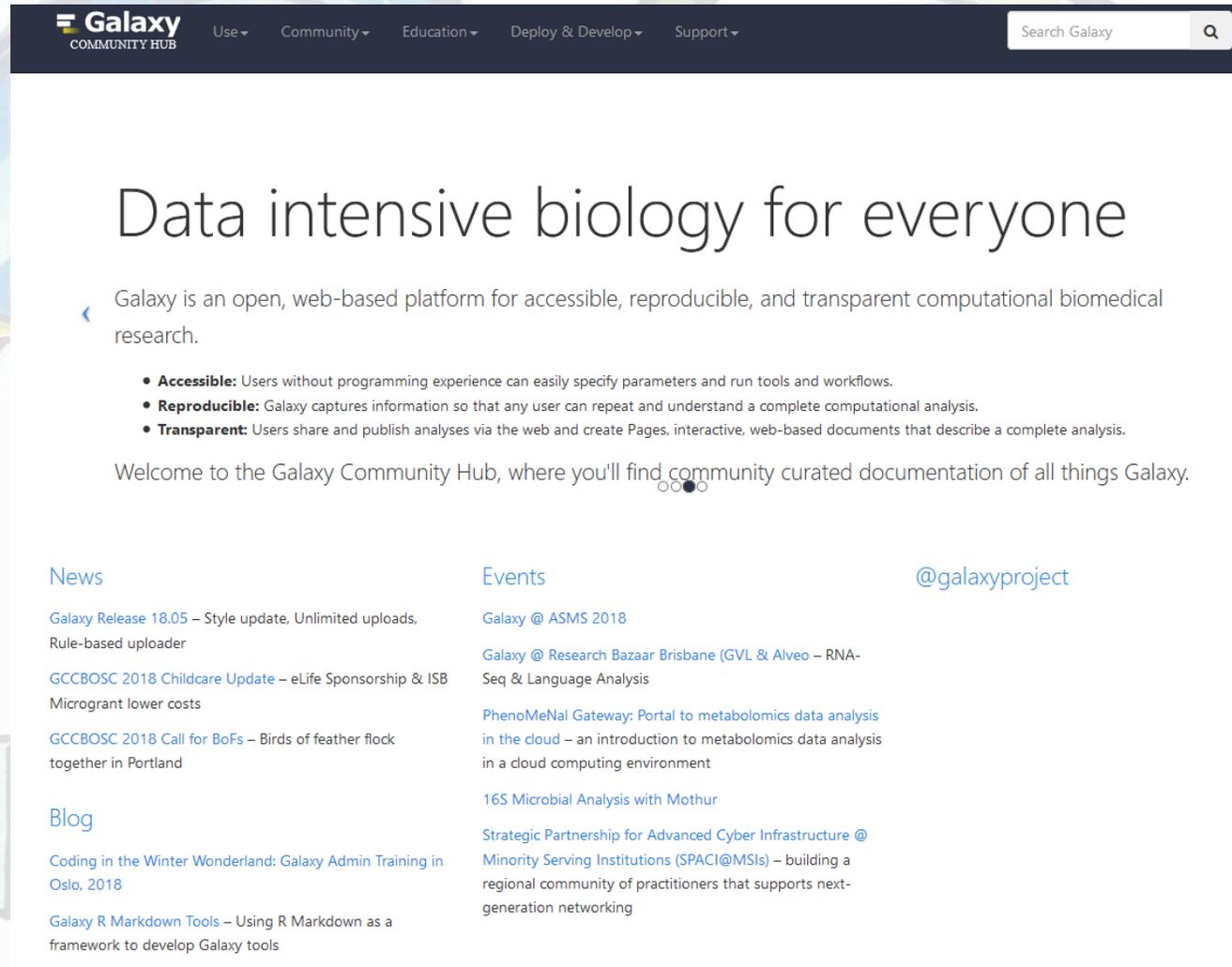


**3rd Course on Bioinformatics tools for
Next Generation Sequencing data mining:
use of bioinformatics tools for typing pathogenic *E. coli***

**IT infrastructure and user interface:
The Galaxy architecture and
ARIES cluster**

Arnold Knijn
EURL VTEC - SANSPV - ISS

<https://galaxyproject.org/>



The screenshot shows the Galaxy Community Hub website. At the top is a dark navigation bar with the Galaxy logo and menu items: Use, Community, Education, Deploy & Develop, and Support. A search bar is on the right. The main content area features a large heading 'Data intensive biology for everyone' and a sub-heading 'Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research.' Below this is a list of key features: Accessible, Reproducible, and Transparent. A welcome message follows. The bottom section is divided into three columns: News, Events, and @galaxyproject, each with several links to recent updates and events.

Galaxy
COMMUNITY HUB

Use ▾ Community ▾ Education ▾ Deploy & Develop ▾ Support ▾

Search Galaxy

Data intensive biology for everyone

Galaxy is an open, web-based platform for accessible, reproducible, and transparent computational biomedical research.

- **Accessible:** Users without programming experience can easily specify parameters and run tools and workflows.
- **Reproducible:** Galaxy captures information so that any user can repeat and understand a complete computational analysis.
- **Transparent:** Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.

Welcome to the Galaxy Community Hub, where you'll find community curated documentation of all things Galaxy.

News

- [Galaxy Release 18.05](#) – Style update, Unlimited uploads, Rule-based uploader
- [GCCBOSC 2018 Childcare Update](#) – eLife Sponsorship & ISB Microgrant lower costs
- [GCCBOSC 2018 Call for BoFs](#) – Birds of feather flock together in Portland

Events

- [Galaxy @ ASMS 2018](#)
- [Galaxy @ Research Bazaar Brisbane \(GVL & Alveo](#) – RNA-Seq & Language Analysis
- [PhenoMeNal Gateway: Portal to metabolomics data analysis in the cloud](#) – an introduction to metabolomics data analysis in a cloud computing environment
- [16S Microbial Analysis with Mothur](#)
- [Strategic Partnership for Advanced Cyber Infrastructure @ Minority Serving Institutions \(SPACI@MSIs\)](#) – building a regional community of practitioners that supports next-generation networking

@galaxyproject

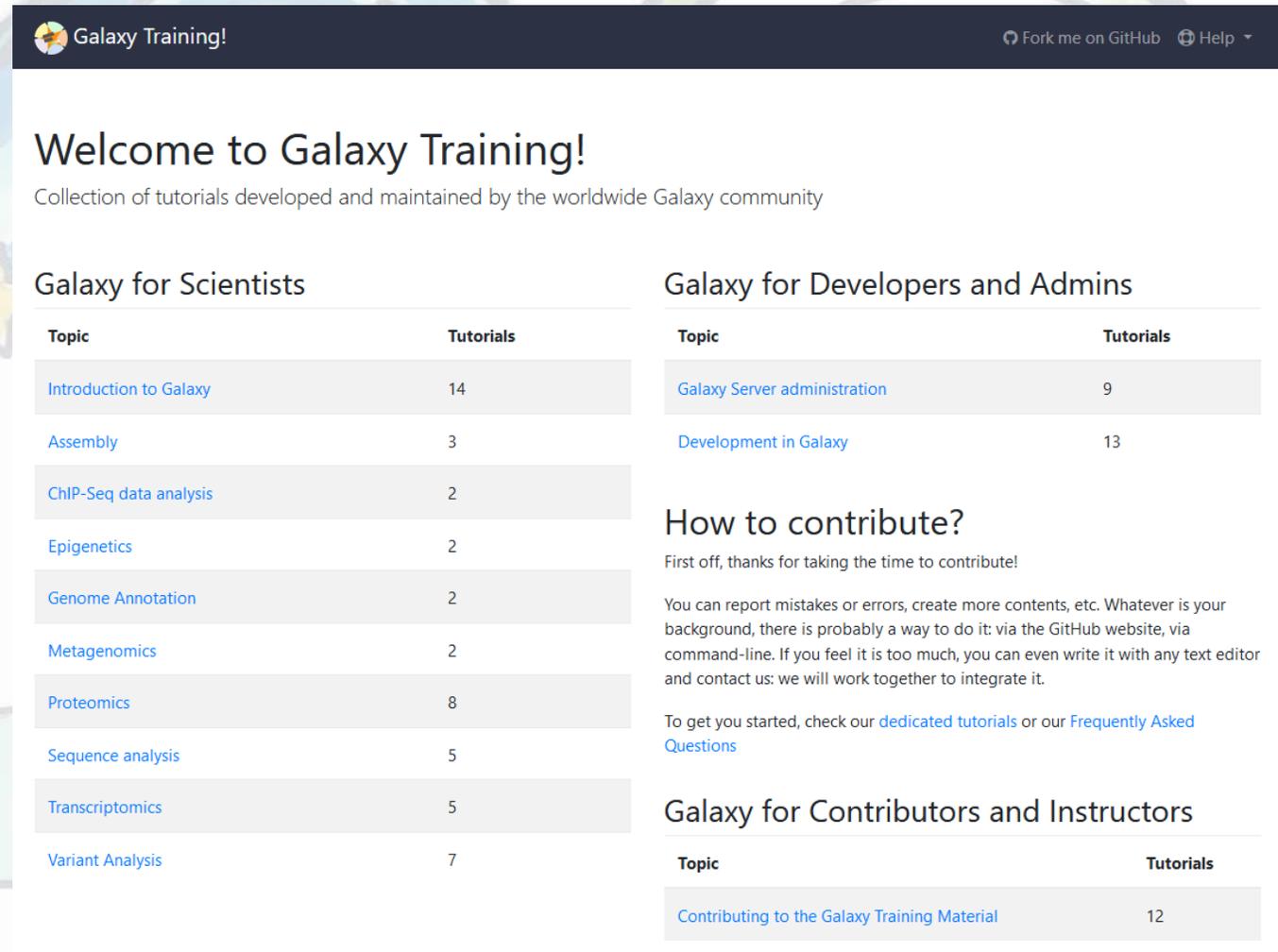
Blog

- [Coding in the Winter Wonderland: Galaxy Admin Training in Oslo, 2018](#)
- [Galaxy R Markdown Tools](#) – Using R Markdown as a framework to develop Galaxy tools

Galaxy core values

- **Accessibility**
 - Users without programming experience can easily upload/retrieve data, run complex tools and workflows, and visualize data
- **Reproducibility**
 - Galaxy captures information so that any user can understand and repeat a complete computational analysis
- **Transparency**
 - Users can share or publish their analyses (histories, workflows, visualizations)

You're not alone: the Galaxy community



Galaxy Training! [Fork me on GitHub](#) [Help](#)

Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

Galaxy for Scientists

Topic	Tutorials
Introduction to Galaxy	14
Assembly	3
ChIP-Seq data analysis	2
Epigenetics	2
Genome Annotation	2
Metagenomics	2
Proteomics	8
Sequence analysis	5
Transcriptomics	5
Variant Analysis	7

Galaxy for Developers and Admins

Topic	Tutorials
Galaxy Server administration	9
Development in Galaxy	13

How to contribute?

First off, thanks for taking the time to contribute!

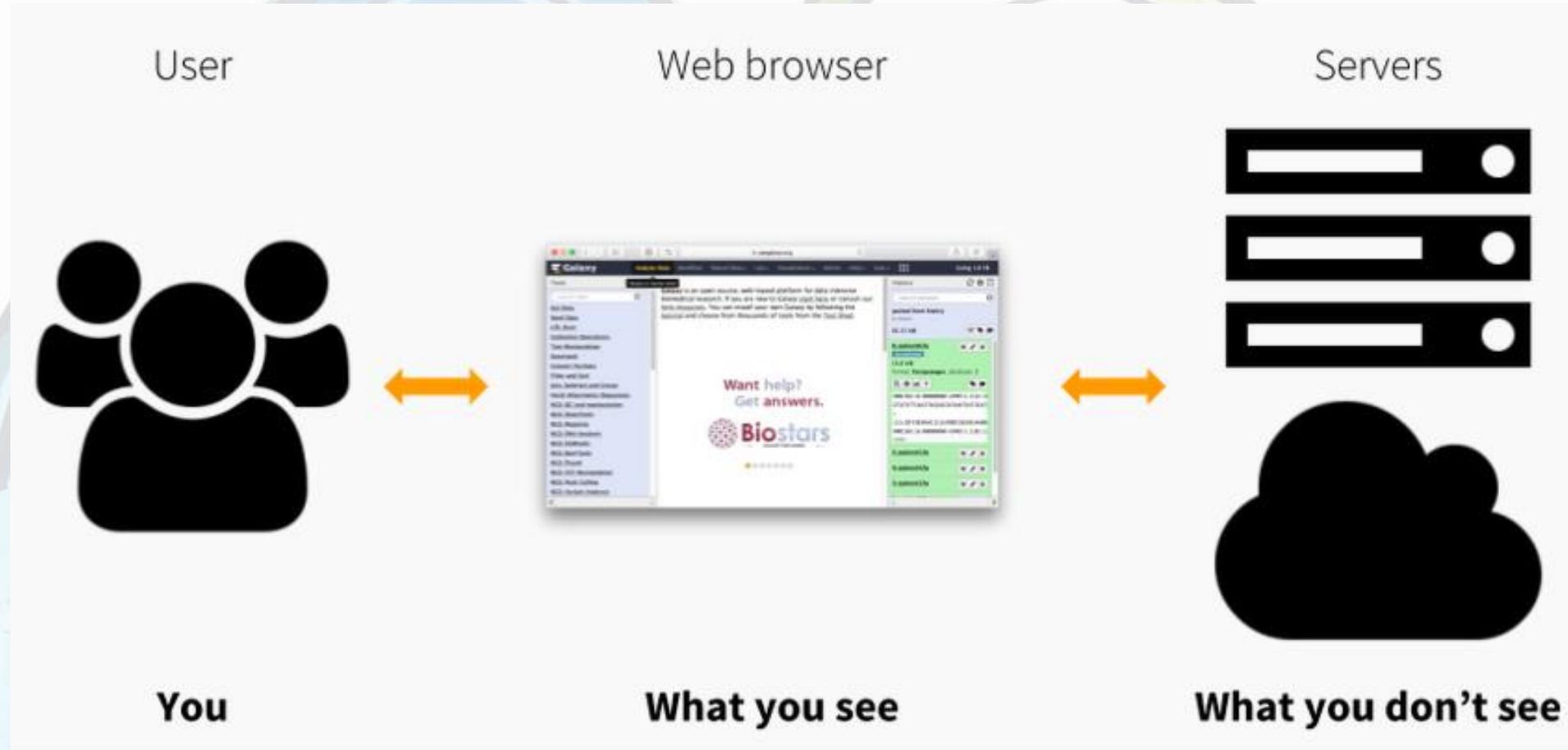
You can report mistakes or errors, create more contents, etc. Whatever is your background, there is probably a way to do it: via the GitHub website, via command-line. If you feel it is too much, you can even write it with any text editor and contact us: we will work together to integrate it.

To get you started, check our [dedicated tutorials](#) or our [Frequently Asked Questions](#)

Galaxy for Contributors and Instructors

Topic	Tutorials
Contributing to the Galaxy Training Material	12

Client-server architecture



Galaxy user interface

Galaxy / ARIES - ISS Analyze Data Workflow Shared Data Visualization Admin Help User Using 1.6 GB

Tools

search tools

--- COMMON TOOLS ---

- Get Data
- Send Data
- Lift-Over
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Statistics
- Graph/Display Data
- GraPhlAn

---HREVAP TOOLS---

- HReVAP

---NGS TOOLS---

- NGS: Assembly
- NCBI Blast
- Manipulation
- kSNP3
- Gene Annotation
- FASTA manipulation
- NGS: Mapping
- NGS: SAM Tools
- NGS: QC and manipulation

Istituto Superiore di Sanità

ARIES - Advanced Research Infrastructure for Experimentation in Genomics - Galaxy Instance at ISS

Tweets by @ARIES_GENOMICS

Aries Group Retweeted

TGAC @GenomeAnalysis

5 reasons #computing isn't scary & why learning to #code is one of the most useful skills bit.ly/1q5oEJF

Five reasons why computing isn't as scary as you think

07 May

Embed View on Twitter

Please read our [disclaimer](#) before using ARIES.

Galaxy is an open, web-based platform for data intensive biomedical research. The Galaxy team is a part of BX at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University. The Galaxy Project is supported in part by NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, and Emory University.

History

search datasets

Unnamed history
287 shown, 665 deleted, 7 hidden
2.77 GB

- 955: BAM-to-SAM on data 147: converted SAM
- 948: Cut on data 947
- 947: Select on data 942
- 945: Cut on data 942
- 944: Cut on data 942
- 943: Cut on data 942
- 942: Galaxy1128-[BAM-to-SAM on data 1124 converted SAM].sam
- 941: comment on data 194, data 193, and others: heatmap_plain.png
- 940: comment on data 194, data 193, and others: heatmap_percentage.png

Galaxy user interface

MENU

TOOLS

WORKING AREA

DATA

Galaxy / ARIES - ISS Analyze Data Workflow S MENU Admin Help User Using 1.6 GB

Bin Assigner (Galaxy Version 1.0.0) Options

Melting Temperatures Tab

955: BAM-to-SAM on data 147: converted SAM

Column to consider to Assign Bin

Execute

Loads the file with the temperatures; matches the temperatures with reference bins and assigns the corresponding alleles

History

455: Escherichia coli classification of data 454

454: Tm_15_strains_geniTolti_Null.txt

452: Escherichia coli classification of data 451

451: Tm_77_new.txt

450: E77Geni clus data

449 clas DATA

448: E77Geni classification of data 447

447: Tm_77qeni.txt

444: Tm_15_new.txt

440: Tm_15_strains_geniTolti.txt

Escherichia coli classification of data 393

438: Escherichia coli

Intuitive and self-documenting

Bin Assigner Home-made tool for bin assignment (Galaxy Version 1.0.0) Options

Melting Temperatures Tab

9492: MeltingTemps_bins.tab

Column to consider to Assign Bin

Execute

Loads the file with the temperatures; matches the temperatures with reference bins and assigns the corresponding alleles

Citations Show BibTeX

Michelacci, Valeria and Orsini, Massimiliano and Knijn, Arnold and Delannoy, Sabine and Fach, Patrick and Caprioli, Alfredo and Morabito, Stefano (2016). Development of a High Resolution Virulence Allelic Profiling (HReVAP) Approach Based on the Accessory Genome of Escherichia coli to Characterize Shiga-Toxin Producing E. coli (STEC). In *Frontiers in Microbiology*, 7.
[doi:10.3389/fmicb.2016.00202][Link]

Home-made tools

```
<tool id="binassigner" name="Bin Assigner">
  <description>Home-made tool for bin assignment</description>
  <requirements>
    <requirement></requirement>
  </requirements>
  <command interpreter="python">
    BinAssigner.py -t $tmstab -o $output -c $columnstab > $logfile
  </command>
  <inputs>
    <param name="tmstab" type="data" format="tabular" label="Melting Temperatures Tab"/>
    <param name="columnstab" type="text" format="integer" label="Column to consider to Assign Bin" />
  </inputs>
  <outputs>
    <data format="tabular" name="output" label="Allele Table"/>
    <data format="txt" name="logfile" label="BinAssigner Log File" />
  </outputs>
  <help>
    Loads the file with the temperatures; matches the temperatures with reference bins and assigns
    the corresponding alleles
  </help>
  <citations>
    <citation type="doi">10.3389/fmicb.2016.00202</citation>
  </citations>
</tool>
```


Which Galaxy flavour?

- Public server (90+)



- Own computer



- Appliance Galaxy Edition



- Data Center



- Cloud



ARIES

ARIES

Analysis of genomic data in the field of public health and food safety, with the aim of deploying a comprehensive bioinformatics approach to the study of food-borne zoonoses and infectious diseases at the human and animal interface.

This is a **Domain specific server**.

Links

- [ARIES](#)

Comments

- The web portal is termed ARIES (Advanced Research Infrastructure for Experimentation in genomicS)
- Aims
 - Development of an Information System for the collection of genomic and epidemiological data to enable the Next Generation Sequencing (NGS)-based surveillance of infectious epidemics, foodborne outbreaks and diseases at the animal-human interface.
 - Development of analytical pipelines enabling harmonized, real time multi-genome comparisons, to improve the detection of clusters of cases of infections and allowing the global bio-tracing of pathogens.
 - Development of metagenomics models for the culture-independent detection and typing of pathogens and the study of their interactions with the microbiota in human and animal samples and in the vehicles of infections.

User Support

- Email support: aries@iss.it
- The system can be rebooted any time for service purposes, although generally a warning will be published with some advance on the ARIES site and through the [@ARIES_GENOMICS](#) Twitter account.
- No backup of data is performed on our servers

Quotas

- An account is required. Contact aries@iss.it to request an account.

Citation(s)

- [Advanced Research Infrastructure for Experimentation in Genomics \(ARIES\): A concept of a bioinformatics framework for the analysis of genomic data from zoonotic agents](#), Valeria Michelacci, Arnold Knijn, Massimiliano Orsini, Stefano Morabito, [4th Med-Vet-Net Association International Conference](#).
- [ARIES tagged publications](#) in [Galaxy Publication library](#).

Sponsor(s)

- Set up and maintained at the [Istituto Superiore di Sanità (ISS Rome)] by a collaboration between the European Union [Reference Laboratory for VTEC](#) and the Data Management Sector.



ISS IT infrastructure

Blade server technology



Storage



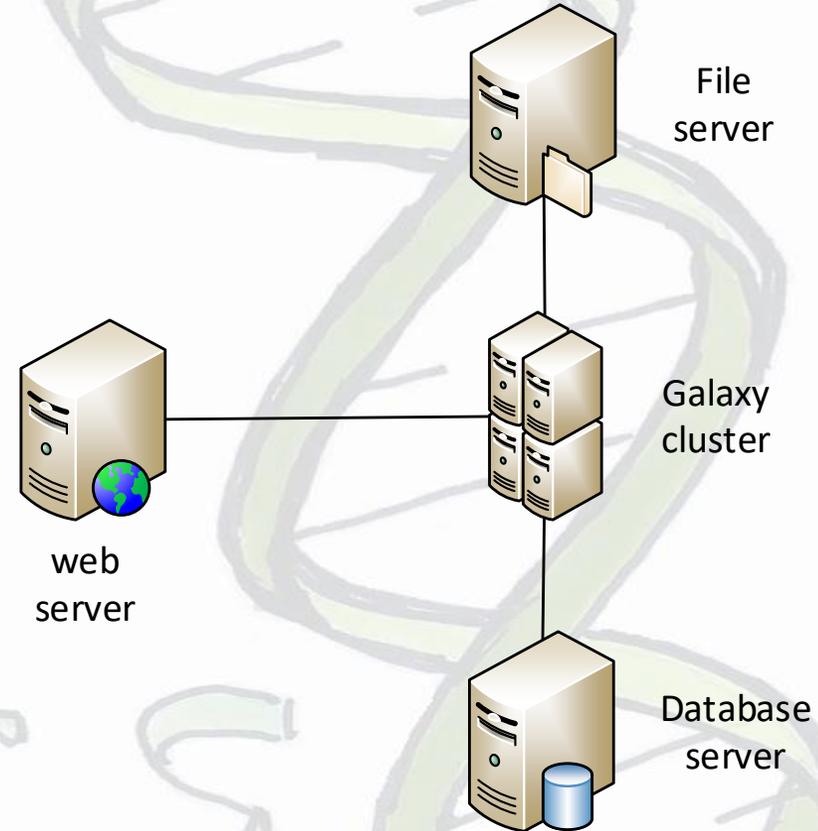
Processing



Networking

Galaxy components

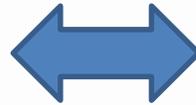
- Web server
- Job runner
- File server
- Database server



Default vs production installation

All-in-one (default)

- ✗ Internal file system
- ✗ SQLite
- ✗ Built-in HTTP server for all tasks
- ✗ Local job runner
- ✗ Single process
- ✓ Simplest error-proof configuration

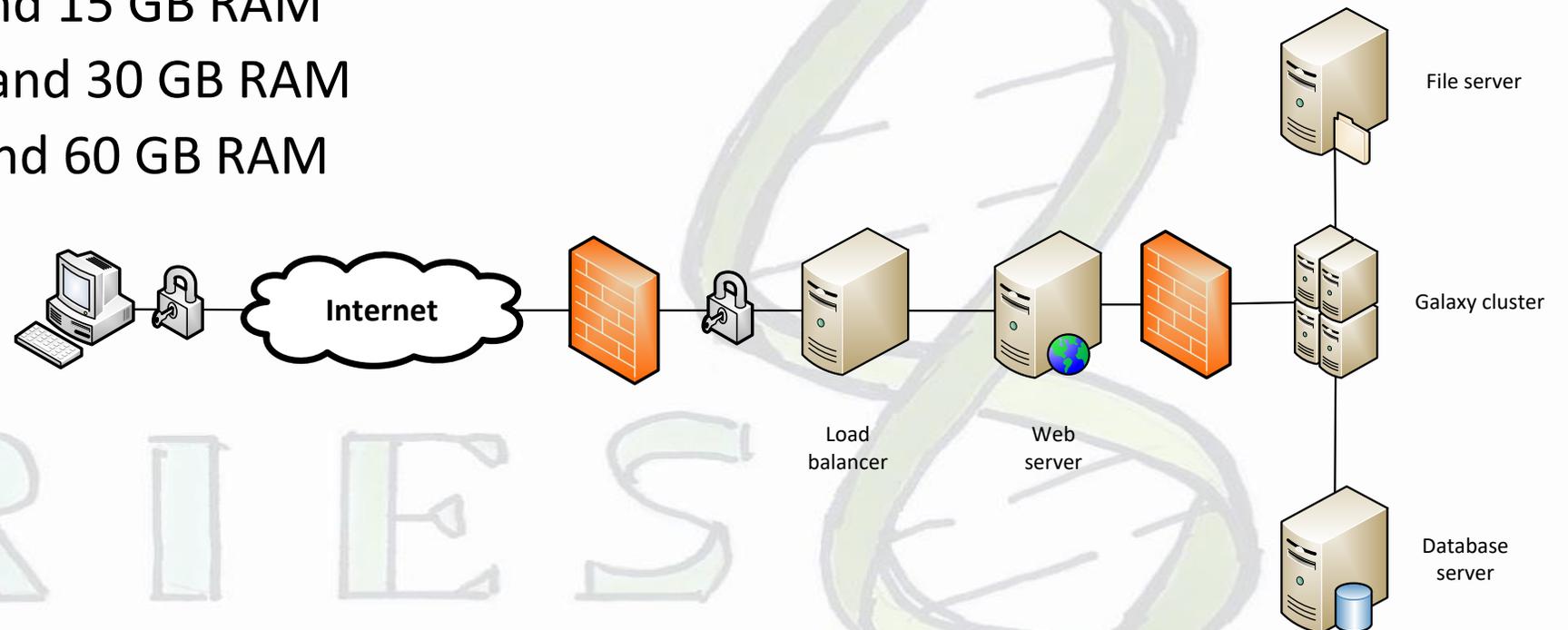


Production (scalable)

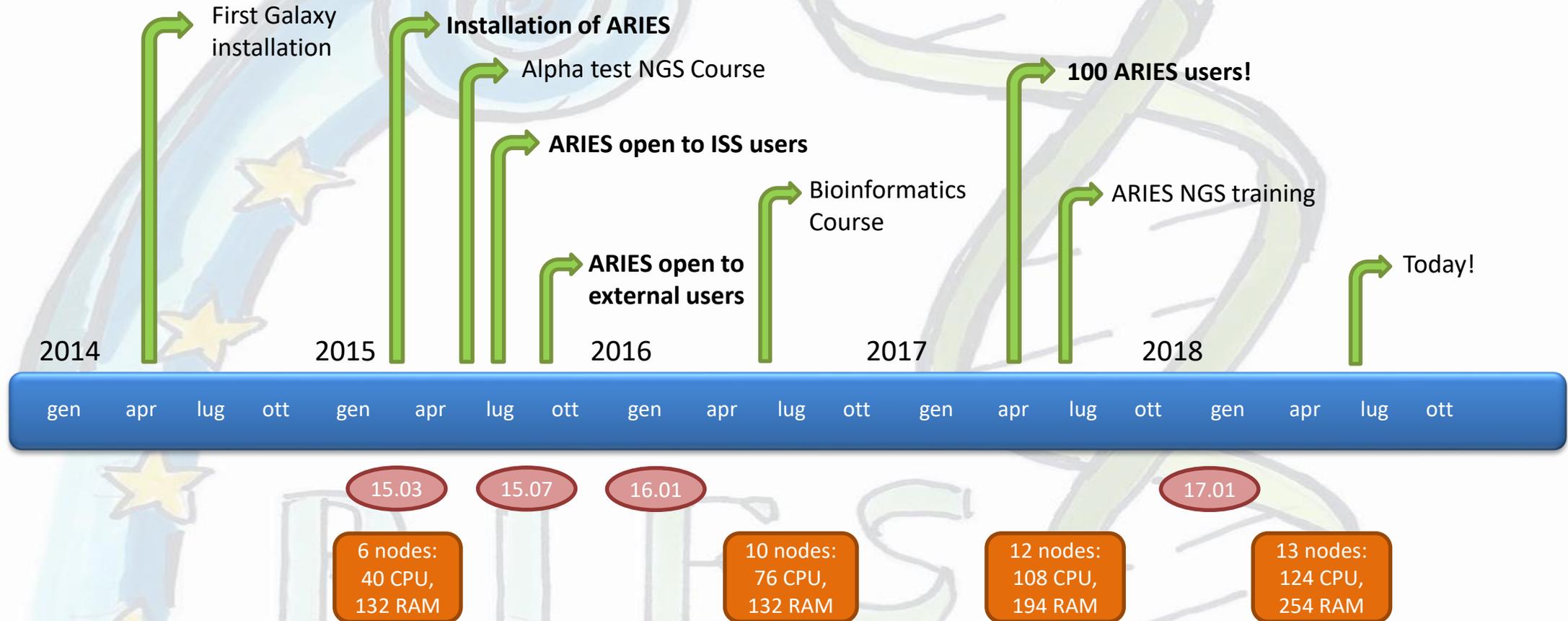
- ✓ External file server
- ✓ External database
- ✓ External HTTPS server for many tasks
- ✓ Cluster job runner
- ✓ Multi process
- ✗ More complex configuration

ARIES cluster

- Cluster is made up of 13 nodes:
- Master node: 8 cores and 8 GB RAM
- 1 node: 4 cores and 4 GB RAM
- 8 nodes: 8 cores and 15 GB RAM
- 2 nodes: 16 cores and 30 GB RAM
- 1 node: 16 cores and 60 GB RAM



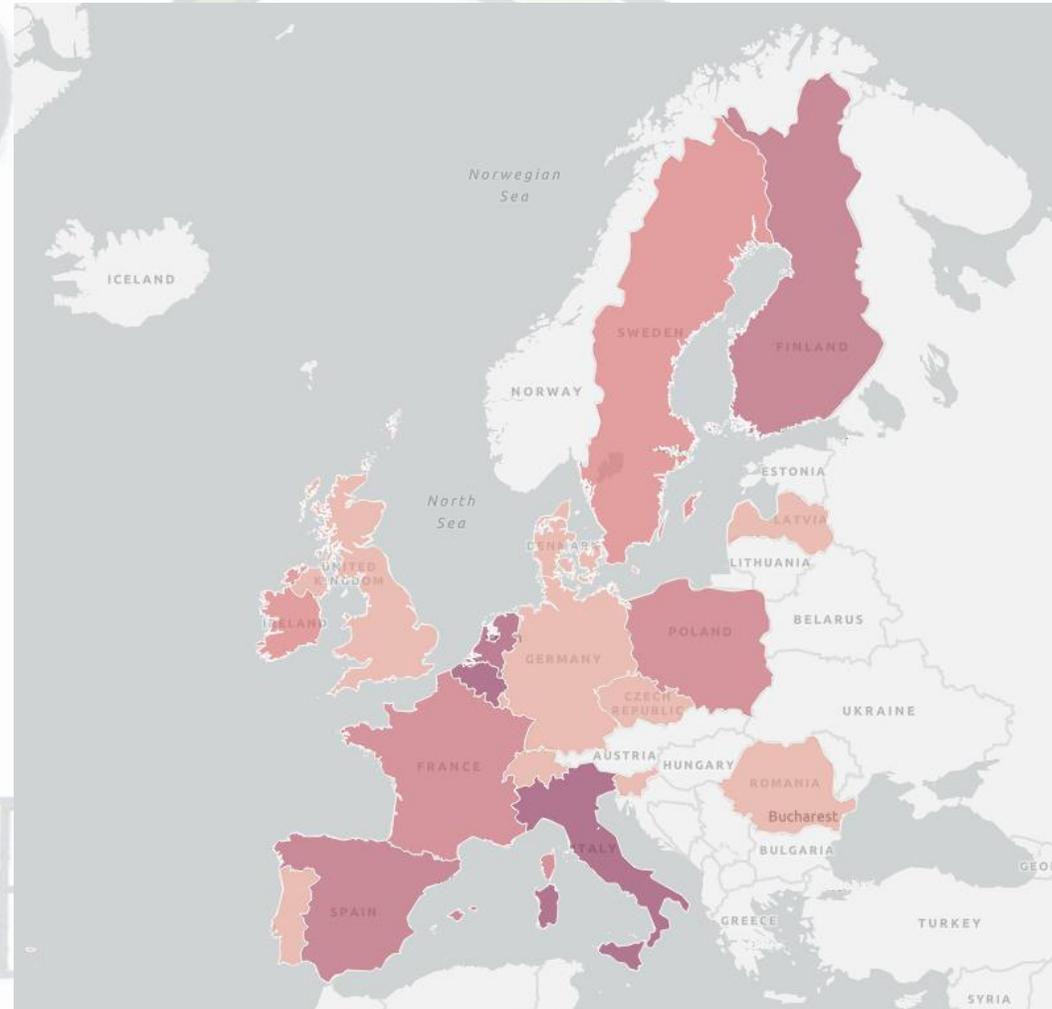
Galaxy in ISS



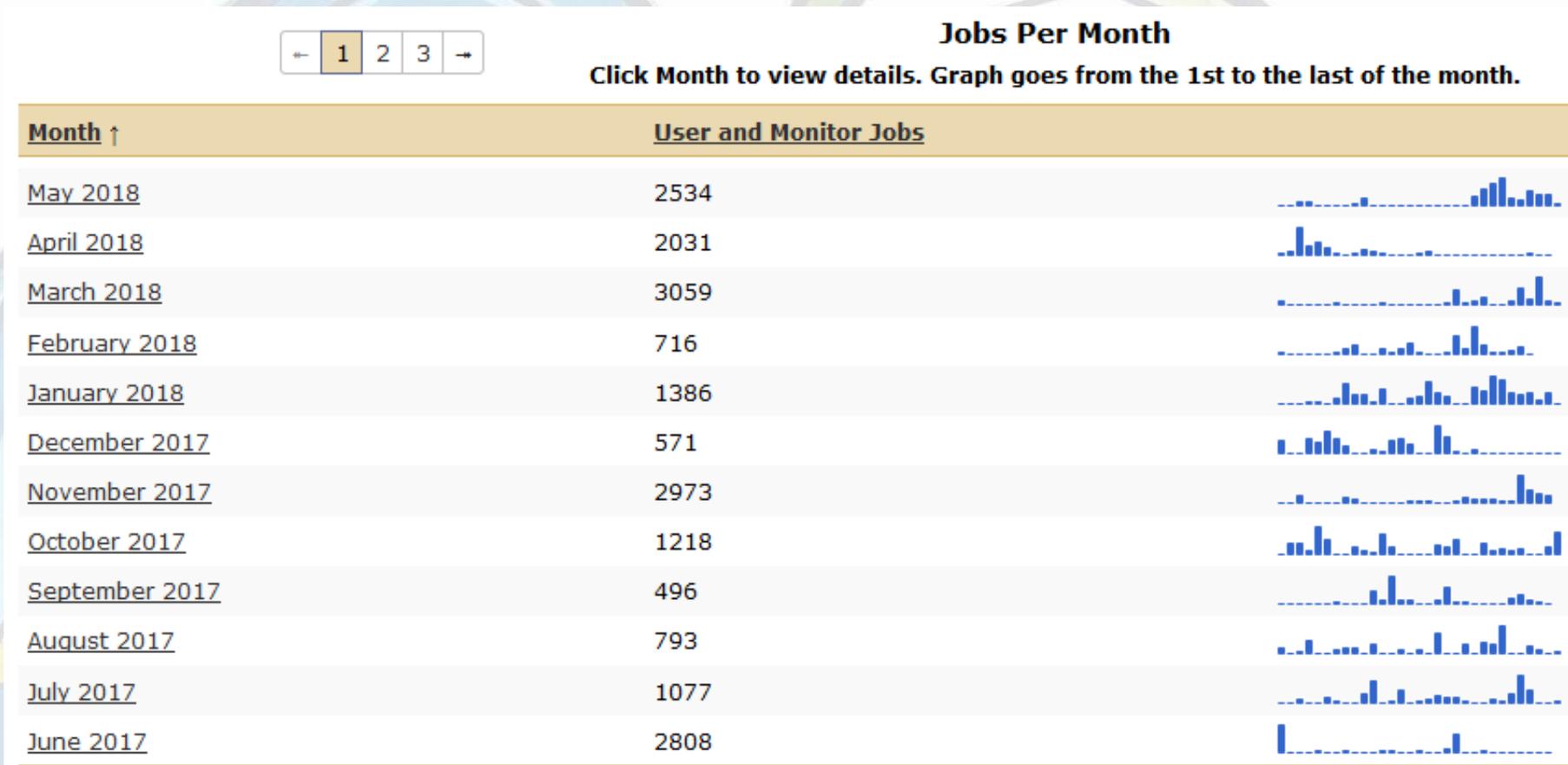
ARIES users

Currently 105 users:

- 4 Belgium
- 1 Czech Republic
- 1 Denmark
- 4 Finland
- 3 France
- 1 Germany
- 2 Ireland
- 51 Italy (37 ISS)
- 1 Latvia
- 1 Luxembourg
- 6 The Netherlands
- 3 Poland
- 1 Portugal
- 1 Romania
- 1 Slovenia
- 2 Spain
- 2 Sweden
- 1 Switzerland
- 1 United Kingdom
- 2 Argentina
- 1 Australia
- 3 China
- 1 India
- 2 Iran
- 3 USA



ARIES use



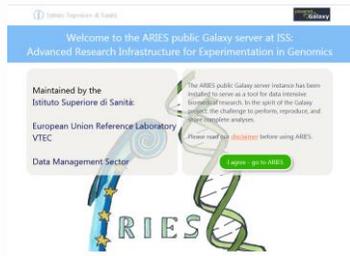
Questions?

ARIES group

- Stefano Morabito
- Valeria Michelacci
- Arnold Knijn

- Massimiliano Orsini

- IZS delle Venezie



<https://w3.iss.it/site/aries>



aries@iss.it



[@ARIES GENOMICS](https://twitter.com/ARIES_GENOMICS)