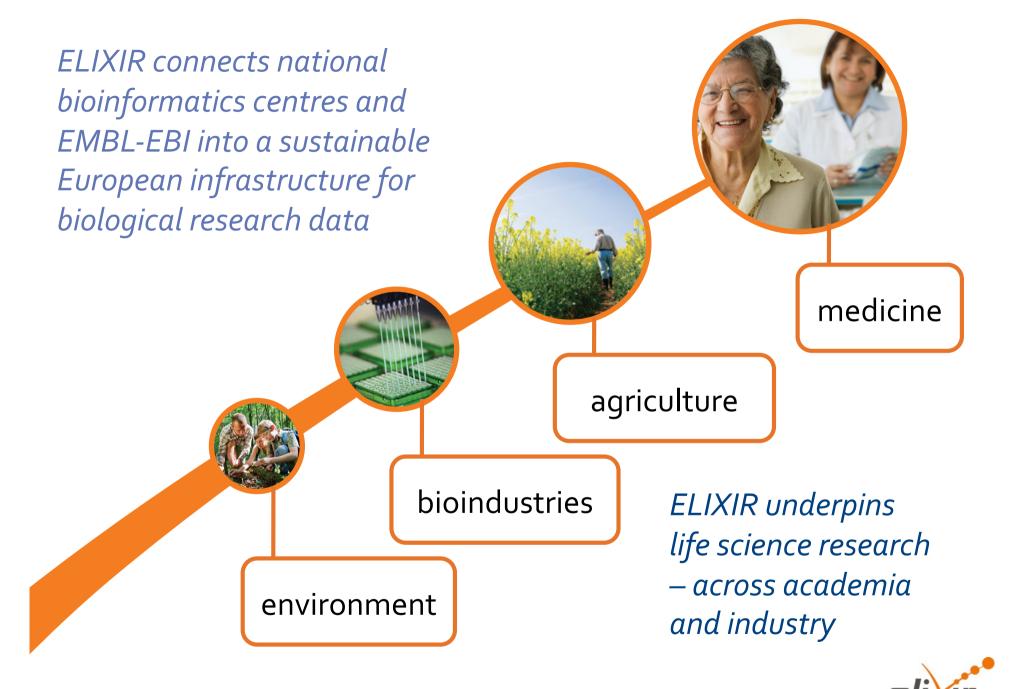
Enabling a federated environment to support biomedical research

Gianmauro Cuccuru – CRS4





ELIXIR: the European Research Infrastructure for

biological data

 ELIXIR connects national infrastructures and EMBL-EBI

16 Member states + EMBL

•Belgium, Czech Republic, Estonia, Denmark, Finland, France, Israel, Italy, Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, Switzerland, UK

 2 countries working towards membership

· Greece, Ireland



Search

ELIXIR

A distributed infrastructure for life-science information

Green light for Italy

Thursday, January 7, 2016

Italy is the latest country to sign the ELIXIR Consortium Agreement and join ELIXIR as Member, following the ELIXIR Board's approval of Italy's application.

The ELIXIR Node in Italy is organised as a Joint Research Unit, coordinated by the National Research Council (CNR). It brings together 13 partners, including seven Italian universities and four High Performance Computing (HPC) centres.



The Italian bioinformatics community has a strong track record in human, animal and plant genomics and has been involved in several genome projects (for example the Sardinians genome project, the

domestic horse and the grape and peach genomes). Other areas of focus for the Node include metagenomics, epigenomics, proteomics and bioinformatics training. ELIXIR will in turn also benefit from the experience and expertise of Italy's four High Performance Computing centres in managing HPC services in various scientific fields.

Prof. Graziano Pesole from the University of Bari and the ELIXIR Italy Head of Node said: "Our goal is to drive the development of bioinformatics resources in Italy and integrate them with our European partners across ELIXIR. We are looking forward to sharing our expertise and resources with ELIXIR partners and bringing them closer to the European research community."

Italy is the 16th ELIXIR Member, and most recent country to join following France, Spain, and Belgium.

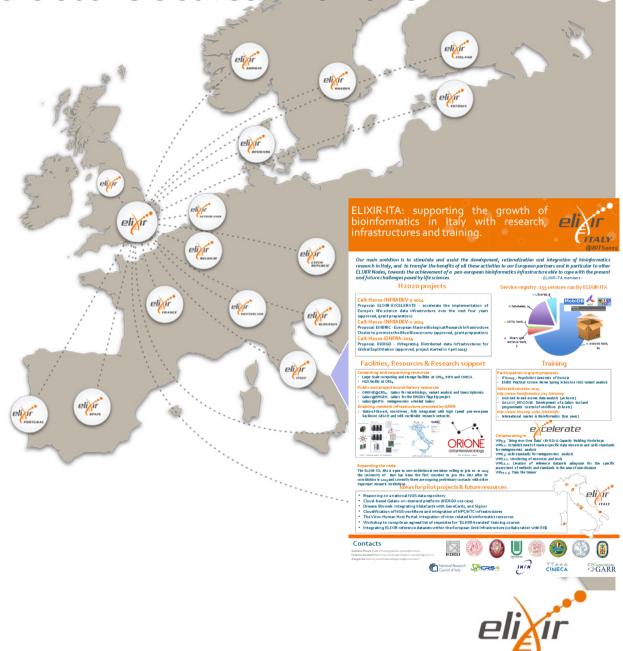


A distributed infrastructure scales with the challenges

 ELIXIR deliver services through national ELIXIR Nodes

 ELIXIR Nodes build local bioinformatics capacity throughout Europe

 ELIXIR Nodes build on national strengths and priorities



ELIXIR Infrastructure

Data

Sustain core data resources

Tools

Services & connectors to drive access and exploitation

Compute

Access, Exchange & Compute on sensitive data

Standards

Integration and interoperability of data and services.

Training

Professional skills for managing and exploiting data



ELIXIR-ITA





Cerca ... Q

HOME STRUTTURA

SERVIZI

EVENTI

NEWS

TRAINING

PROGETTI

CONTATTI

LOGIN



IIB

ELIXIR è l'infrastruttura di ricerca europea (www.elixir-europe.org) per i dati biologici che ha come obiettivo primario quello di supportare la ricerca nel campo delle "scienze della vita" e le attività traslazionali per la medicina, l'ambiente, le industrie biotecnologiche e la società.

ELIXIR è costituita come una infrastruttura distribuita su diversi Nodi, uno per Stato membro, ospitati da centri di eccellenza localizzati in tutta Europa (i "Nodi ELIXIR"). I

Nodi sono collegati ad un hub centrale ('ELIXIR-Hub') che è localizzato ad Hinxton (Regno Unito) presso il "Wellcome Genome Campus" dove è anche ospitato l'European Bioinformatics Institute dell'EMBL (EMBL-EBI).

Il nodo italiano di ELIXIRè coordinato dal CNR e attualmente comprende 17 partner che includono istituti di ricerca, università e istituzioni tecnologiche.

Il Nodo italiano di ELIXIR, che ha l'ambizione di costituire un Istituto Italiano di Bioinformatica (IIB) distribuito su più centri, si propone di aggregare tutti i ricercatori Italiani che operano nel campo della Bioinformatica, favorendo lo scambio e lo sviluppo di competenze, di mettere a sistema le varie risorse bioinformatiche già riconosciute in ambito internazionale e pubblicamente disponibili, e di contribuire alla loro integrazione in seno all'infrastruttura europea.

Un altro obiettivo primario di ELIXIR-ITA è quello di organizzare attività di training, sia di base che avanzato, nei diversi settori applicativi della Bioinformatica al fine di favorire la formazione di giovani bioinformatici, la cui domanda è in fortissima crescita in ambito nazionale ed internazionale.



MARTED

training IIB 09:0

The IIB/ELIXIR-ITA Training Programme

-

Courses

Events&Collaborations

Trainers

Contacts

NEWS >

Seminario di Bioinformatica Strutturale - Prof.

Università di Roma La Sapienza 24/03/2016



Martedi 30 marzo alle ore 14.00 prof. Torsten Schwede dell'Università di Basilea&Swiss Institute of Bioinformatics terrà un seminario dal titolo: "Using evolutionary information for

EVENTI >

BITS2016

15/06/2016 - 17/06/2016 Giorno intero



13° Meeting Annuale della Società di Bioinformatica Italiana 15-17 Giugno 2016, Università di Salerno, Italia La Società di Bioinformatica Italiana, insieme all'Università degli Studi di

TRAININ

Workshops

Università di Salemo 14/06/2016 09:00



2016

1. High Performance Molecular Dynamics

CINECA, via dei Tizii 6, Rome - 5-7 April 2016

This course is designed for those users who wish run classical molecular dynamics programs such as GROMACS and NAMD on modern supercomputers. By understanding better the HPC infrastructures and the algorithms used to exploit them, the aim is to give researchers the tools to run simulations in the most efficient way possible on current and future supercomputers.

At least half the course will be devoted to practical sessions where students will be able to prepare and run GROMACS or NAMD examples on the supercomputers of Cineca.

2. NGS for evolutionary biologists: from basic scripting to variant calling

Biblioteca, Consiglio Nazionale delle Ricerche, via P. Castellino 111, 80131 Naples - May 2-6, 2016

This course will provide an introduction to next generation sequencing platforms, data analysis and tools for data quality control, including alignment to a reference sequence, data handling and visualisation, and variant calling and filtering (single nucleotide polymorphisms and structural variants). The course will be delivered using a mixture of lectures and computer based hands-on practical sessions, including group projects to be completed by the participants using the knowledge gained at the course. Projects will cover the general topics of population structure and admixture, demographic changes and natural selection.

3. RNA-seg data analysis workshop

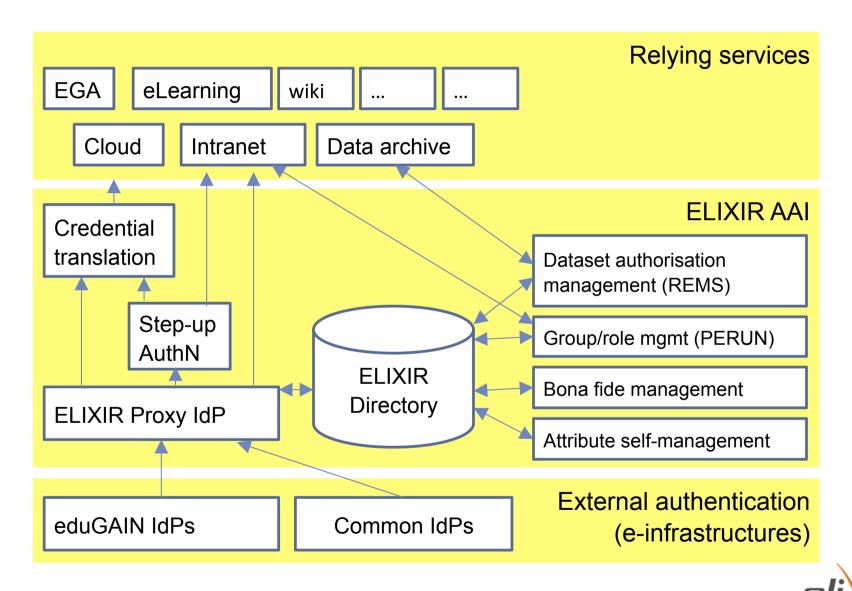
University of Naples Federico II, Naples - June 6-10, 2016

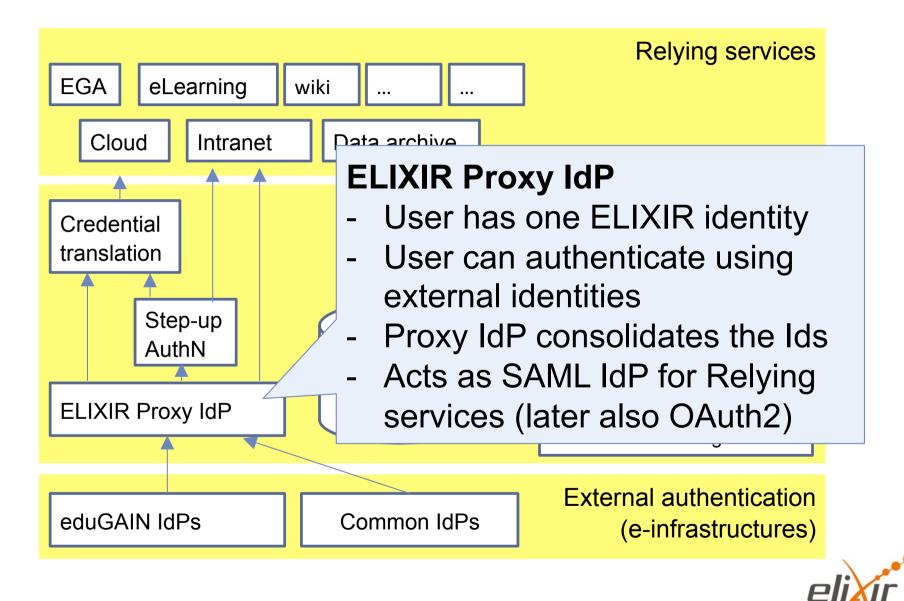
This course will provide an introduction on the application of the RNA-seq technology to the study of non-model organism transcriptomes and it will be centered on the Trinity software suite for the de novo

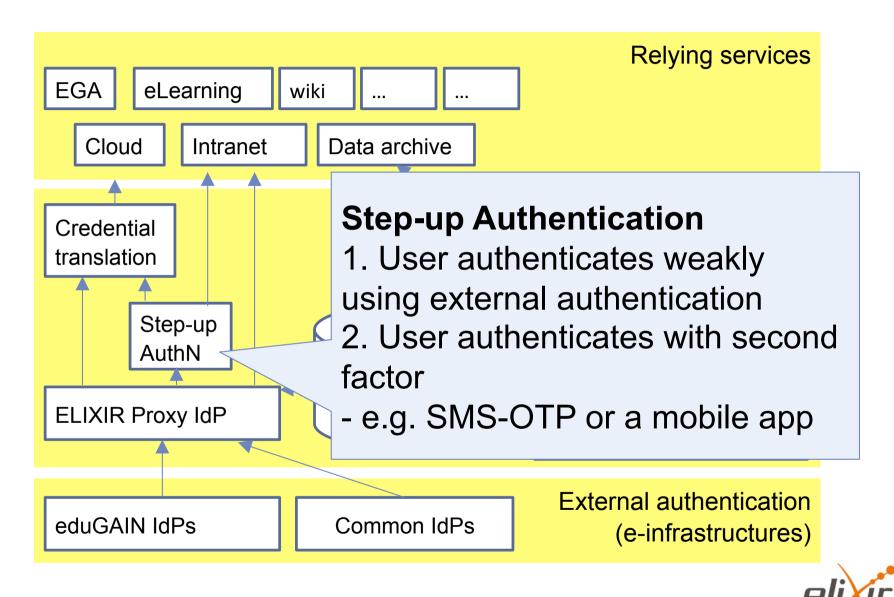
ELIXIR AAI history – where we are now

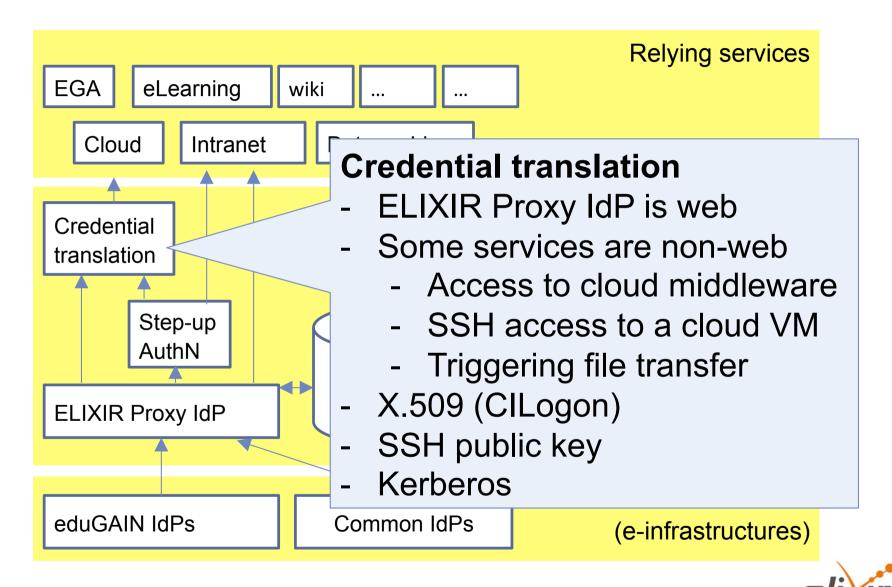
- Use case gathering -- Autumn 2014
 - https://www.elixir-europe.org/documents/elixir-aai-use-cases
- Requirements and design Spring 2015
 - https://www.elixir-europe.org/documents/aai-requirements-and-design
- Deployment starts Autumn 2015 EXCELERATE WP4.3.1
 - Part of ELIXIR Compute platform
- First release -- August 2016
 - Until that ELIXIR AAI in pilot status
 - Key components up and running already

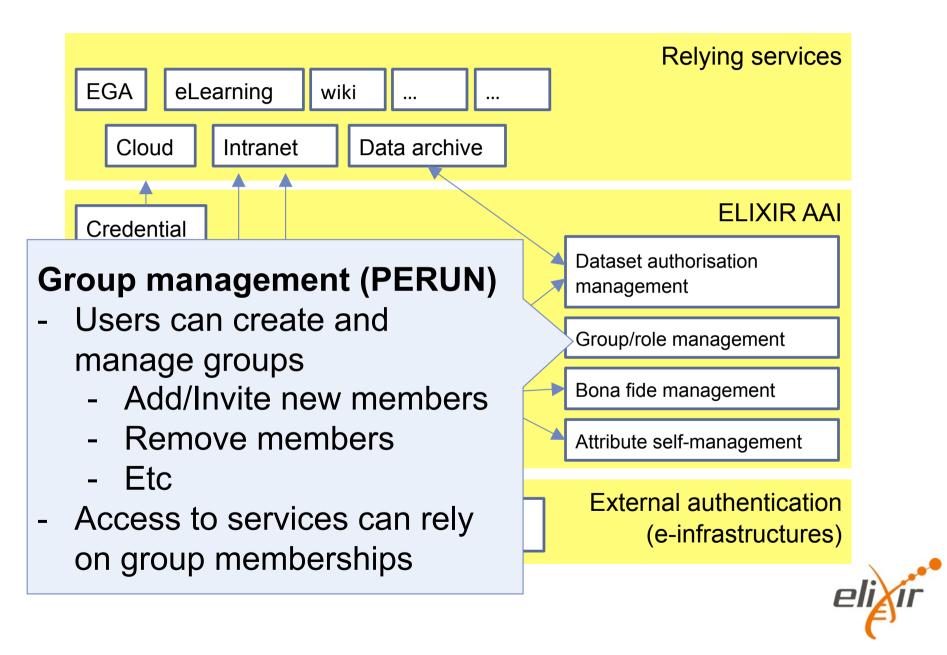




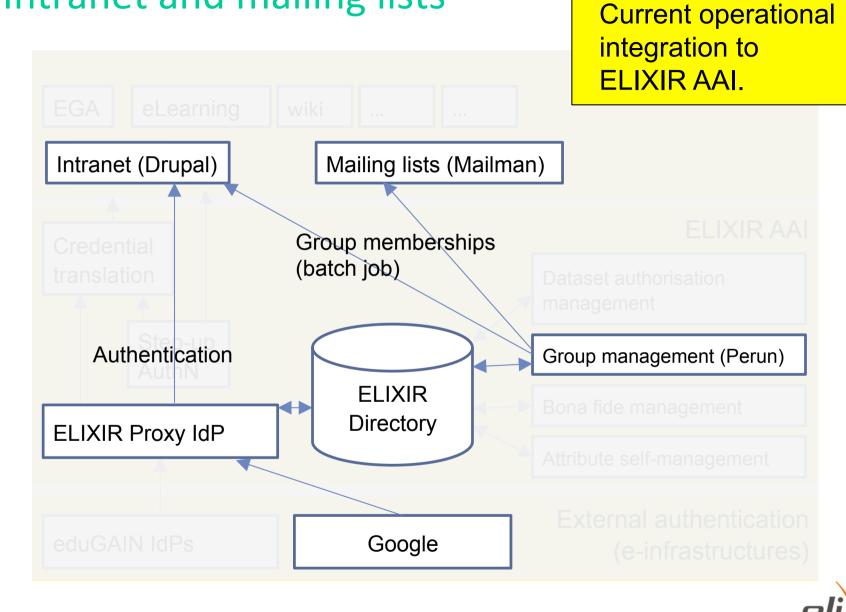








Intranet and mailing lists



Galaxy

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. Galaxy integrates a multitude of different tools by giving them the same "look and feel"

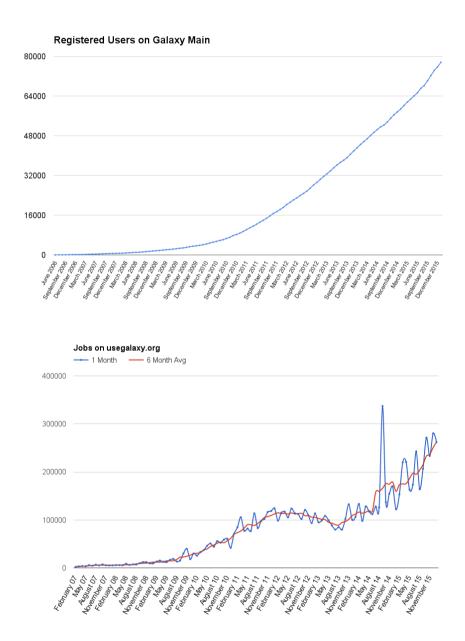
Reproducible: Galaxy captures information so that any user can repeat and understand a complete computational analysis.

Transparent: Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.





Galaxy main usage



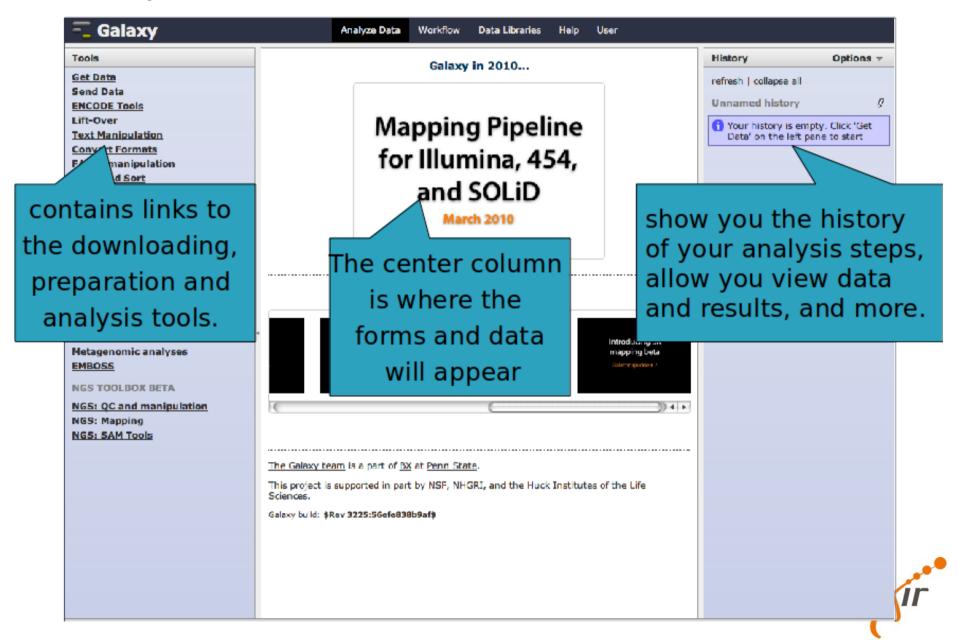
Publicly accessible server

| Date | # Servers |
|---------|-----------|
| 2011/07 | 15 |
| 2012/01 | 21 |
| 2012/07 | 20 |
| 2013/01 | 25 |
| 2013/02 | 27 |
| 2013/07 | 35 |
| 2014/01 | 54 |
| 2014/07 | 60 |
| 2015/01 | 70 |
| 2015/07 | 73 |
| 2016/01 | 84 |

https://wiki.galaxyproject.org/GalaxyProject/Statistics



Galaxy

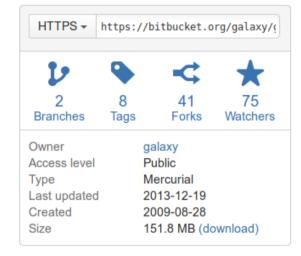


Galaxy



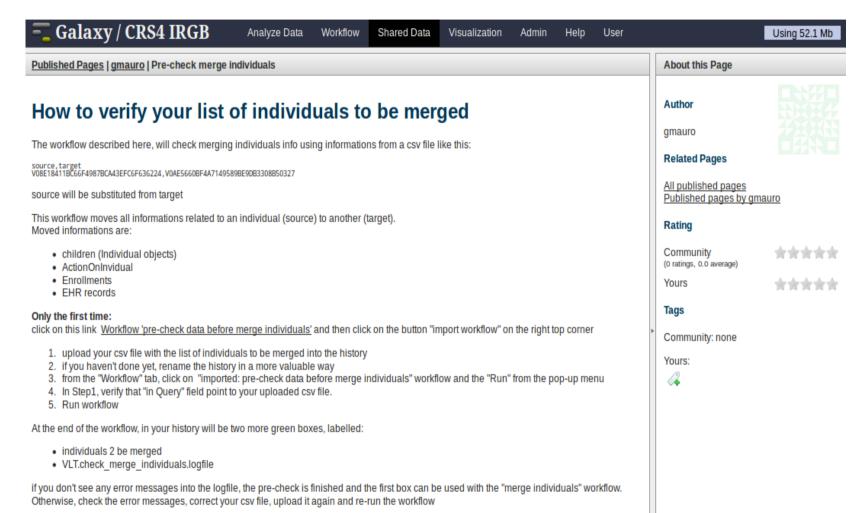








Galaxy - Pages

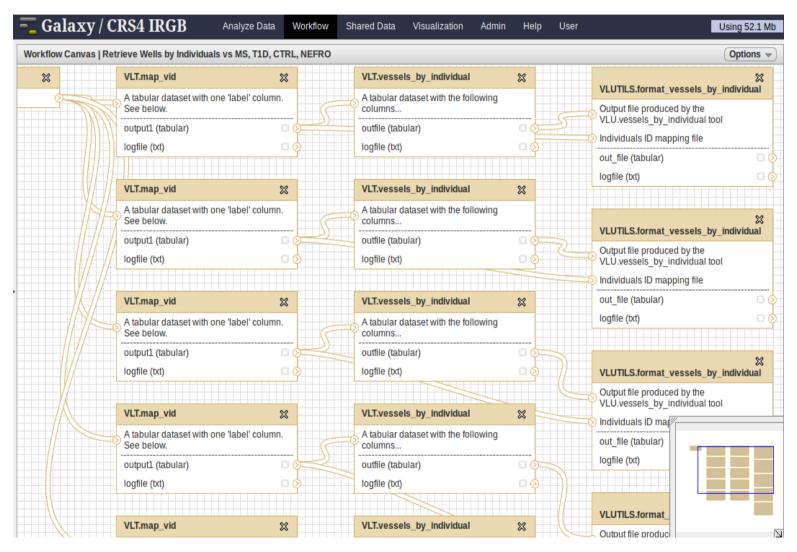




Galaxy - Histories

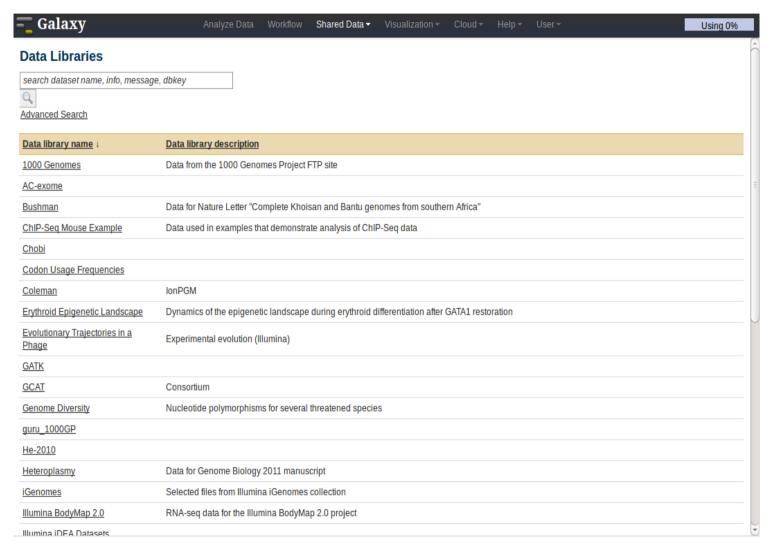
| | | | O | 0 * |
|--|---|---|---|--------------|
| Tools ± | Unknown Tool | | History | 2 \$ |
| search tools | Name: | Map with BWA for Illumina on data 14: mapped reads | search datasets | ε |
| 3541 511 53513 | Created: | Fri Jul 31 13:48:36 2015 (UTC) | | |
| COMMON TOOLS | Filesize: | 21.6 GB | Sex check | |
| Get Data | Dbkey: | hg19 | 19 shown, 14 <u>deleted</u> | |
| IGS: Quality control | Format: | sam | 71.2 GB | S |
| Galaxy 1001 ID. | | bwa_wrapper | 33: | |
| FastQC:Read QC reports using FastQC | Galaxy Tool Version: | 1.2.3 | | |
| • | Tool Version: 0.7.9a-r786 | | VLT.all_enrollments.logfile | |
| IGS: manipulation | Tool Standard Output: | stdout | <u>32:</u> | (4) |
| IGS: Mapping | Tool Standard Error: | stderr | VLT.all enrollments | s.csv |
| ICBI BLAST+ | Tool Exit Code: | 0 | A4. May | |
| TOOLS BY TORKS | API ID: | 824843df326fb7d1 | 31: Map with BWA | |
| TOOLS BY TOPIC | History ID: | 0072e017d28ba2de | for Illumina on data 14: mapped reads | |
| MICROBIOLOGY | UUID: | e9754268-f9d4-4e48-a620-a954b42e62ea | |)E commont |
| IGS: de novo assembly | Full Path: | /home/lgmauro/orione/database/files/000/278/dataset_278399.dat | ~58,000,000 lines, 95 comments format: sam , database: hg19 | |
| ligning and Phylogeny | | python /u/galaxy/orione-tools/toolshed/repos/devteam/bwa_wrappers/ffa8aaa14f7c | Torride. Sain, databa | se. IIg19 |
| Sene annotation | | /bwa_wrappers/bwa_wrapper.pythreads="\${GALAXY_SLOTS:-4}"fileSource="indexed"ref="/SHARE/USERFS/els7/users/biobank/genome | DWA Version: 0.7.94-1780 | |
| delle dilliotation | | /hq19/bwa_index/hq19full/hq19full.fa"input1="/SHARE/USERFS/els7/users | BWA run on paired-e | end data |
| ARIANT CALLING | | /galaxy/orione/user_library_import_dir/andrea.pinna@crs4.it | B 6 2 III | > |
| GS: SNP Effects | | /120702_SN526_0205_AC0BVMACXX/rare_000063_R2_L5.fastq" | | |
| IGS: SAM Tools | | input2="/SHARE/USERFS/els7/users/galaxy/orione/user_library_import_dir | View details AG 3.RNAM | ME 4.POS 5.M |
| IGS: GATK2 | | /andrea.pinna@crs4.it/120702_SN526_0205_AC0BVMACXX | @SQ SN:chrM LN:10 | 5571 |
| CF Tools | Job Command-Line: | /rare_000063_R2_L5.fastq"output="/SHARE/USERFS/els7/users/biobank/galaxy /central/files/000/278/dataset_278399.dat"genAlignType="paired" | @SQ SN:chrl LN:24 | 19250621 |
| | Job Command-Line. | params="full"maxEditDist="0"fracMissingAligns="0.04" | @SQ SN:chr2 LN:24 | 13199373 |
| ED Tools | | maxGapOpens="1"maxGapExtens="-1"disallowLongDel="16" | @SQ SN:chr3 LN:19 | 98022430 |
| reeBayes | | disallowIndel="5"seed="-1"maxEditDistSeed="2"mismatchPenalty="3" | @SQ SN:chr4 LN:19 | 91154276 |
| opy number variation | | gapOpenPenalty="11"gapExtensPenalty="4"suboptAlign="" | @SQ SN:chr5 LN:18 | 30915260 |
| IETAGENOMICS | | noIterSearch="false"outputTopN="3"outputTopNDisc="10"maxInsertSize="500"maxOccurPairing="100000" | | |
| letaPhIAn | | rqid="120702_SN526_0205_AC0BVMACXX.L5.10"rqcn="CRS4"rqds="" | <u>29:</u> | (4) |
| ther metagenomics tools | 1 | rgdt=""rgfo=""rglb="ex12"rgpg=""rgpi="" | VLT.all enrollments | |
| anel metagenomics tools | | rgpl="ILLUMINA"rgpu="None"rgsm="rare_063"suppressHeader="false" | | |
| GS: RNA ANALYSIS | SGE_STDIN_PATH (runtime | /dev/null | 28: | (4) |
| NA-seq | environment variable) | /ucv/muii | VLT.all enrollments | s.csv |
| | SGE_JOB_SPOOL_DIR | | 27: Compute XY | (4) |
| HIP-SEQ | (runtime environment | /var/ge2011_11/execd_spool/entu212/active_jobs/8688189.1 | ratio.logfile | |
| IGS: Peak Calling | variable) | | 20.0 | |
| Notif Tools | UPSTART_EVENTS (runtime environment variable) | runlevel | 26: Compute XY | (4) |
| OTHER TOOLS | FNVIRONMENT (runtime | | ratio.csv | |

Galaxy - Workflows





Galaxy - Libraries





Galaxy – Tool Shed

Next Gen Mappers

Ontology Manipulation

■ Galaxy Tool Shed 3904 valid tools on May 15, 2016 Repositories by Category Search search repository name, description Search for valid tools Search for workflows Name Description Repositories **Valid Galaxy Utilities** Tools for working with assemblies 75 Assembly Tools Tools for analyzing and manipulating ChIP-seq data. 40 Custom datatypes ChIP-seq Repository dependency definitions Tools for combinatorial selection 6 Combinatorial Selections ■ Tool dependency definitions Tools for use in computational chemistry Computational chemistry 24 All Repositories Tools for converting data formats 64 Convert Formats Browse by category Utilities for Managing Galaxy's built-in data cache 32 **Available Actions** Data Managers Login to create a repository 35 Tools for retrieving data from external data sources Data Source Tools for analyzing Epigenetic/Epigenomic datasets 3 Epigenetics 77 Tools for manipulating fasta data Fasta Manipulation Fastq Manipulation Tools for manipulating fastq data 56 Utilities to support Genome-wide association studies 20 Genome-Wide Association Study Tools for operating on genomic intervals 42 Genomic Interval Operations 42 Tools producing images Graphics Utilities to support imaging 1 **Imaging** Tools for use in the study of Metabolomics 25 Metabolomics Tools enabling the study of metagenomes 58 Metagenomics Tools for performing micro-array analysis 8 Micro-array Analysis

Tools for manipulating ontologies

Tools for the analysis and handling of Next Gen sequencing data

105

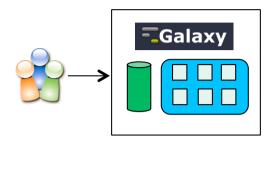
10

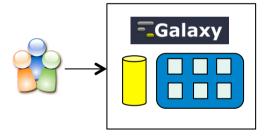
GALAXY@ELIXIR-ITA

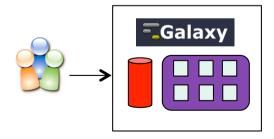
- ORIONE@CRS4: microbiology, variant analysis and transcriptomics.
- Galaxy@EPIGEN: EPIGEN flagship project.
- Galaxy@INFN: metagenomics oriented
- ...



NGalaxy?



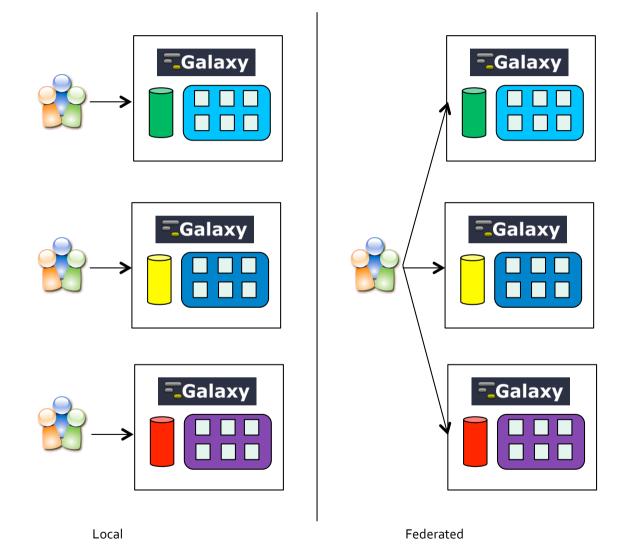




Local

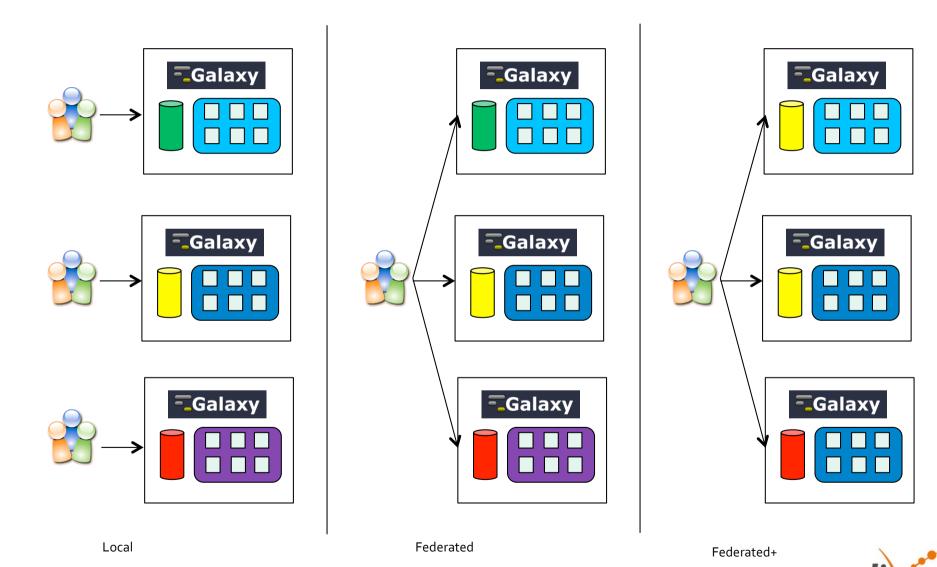


NGalaxy?

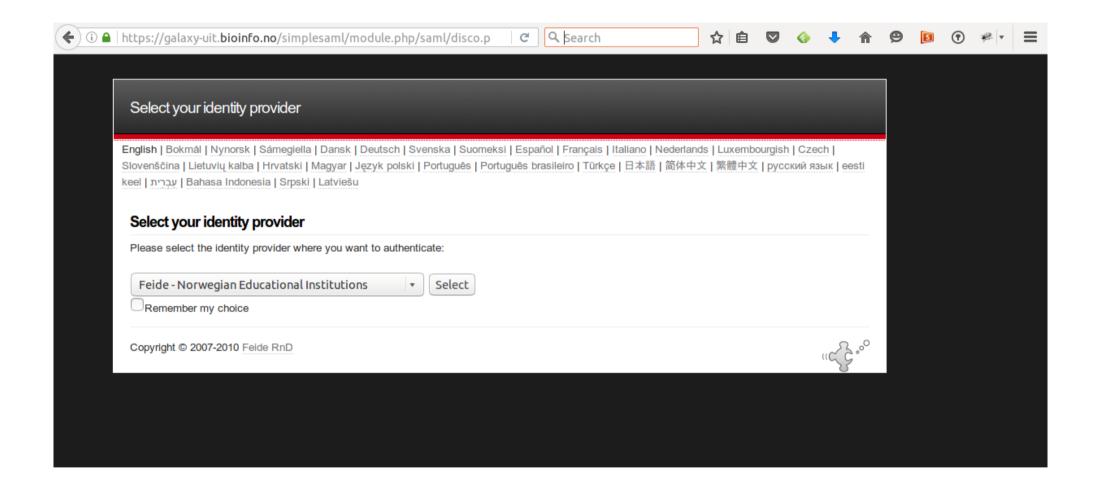




NGalaxy?



Galaxy@ELIXIR-NO





Thank you

