UNIVERSITY OF CAMBRIDGE

Train MALTA

Bioinformatics Training

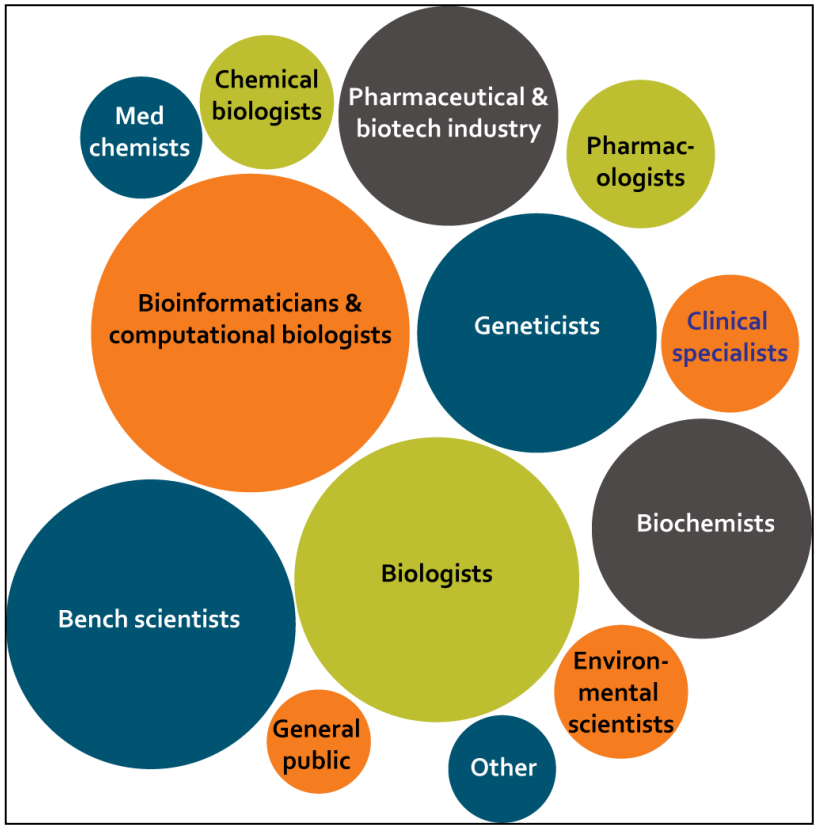
Louisa Bellis, PhD - School of Biological Sciences ***ljb201@cam.ac.uk***

elixir

GOBLET
Global Organisation for Bioinformatics Learning, Education & Training

Why is bioinformatics training important?

- Data analysis is now the major bottleneck to research in the molecular life sciences
- Many biomedical professionals feel under-qualified to make the most of biological data
- There are an estimated 3 million life scientists in Europe alone; >20 million healthcare professionals
- Potentially all of them are producers or consumers of data managed by Europe's biomedical research infrastructures



ELIXIR: a distributed infrastructure for life science

Building a **sustainable European infrastructure for biological information**, supporting life science research and its translation to medicine, agriculture, bioindustries and society

ELIXIR unites Europe's leading life science organisations in **managing and safeguarding the massive amounts of data** being generated every day by publicly-funded research

ELIXIR will **provide the facilities necessary for life science researchers to make the most of our rapidly growing store of information** about living systems, which is the foundation on which our understanding of life is built

ELIXIR "platforms" organisation

- **Data**

  Sustain core data resources

- **Tools**

  Services & connectors to drive access and exploitation

- **Compute**

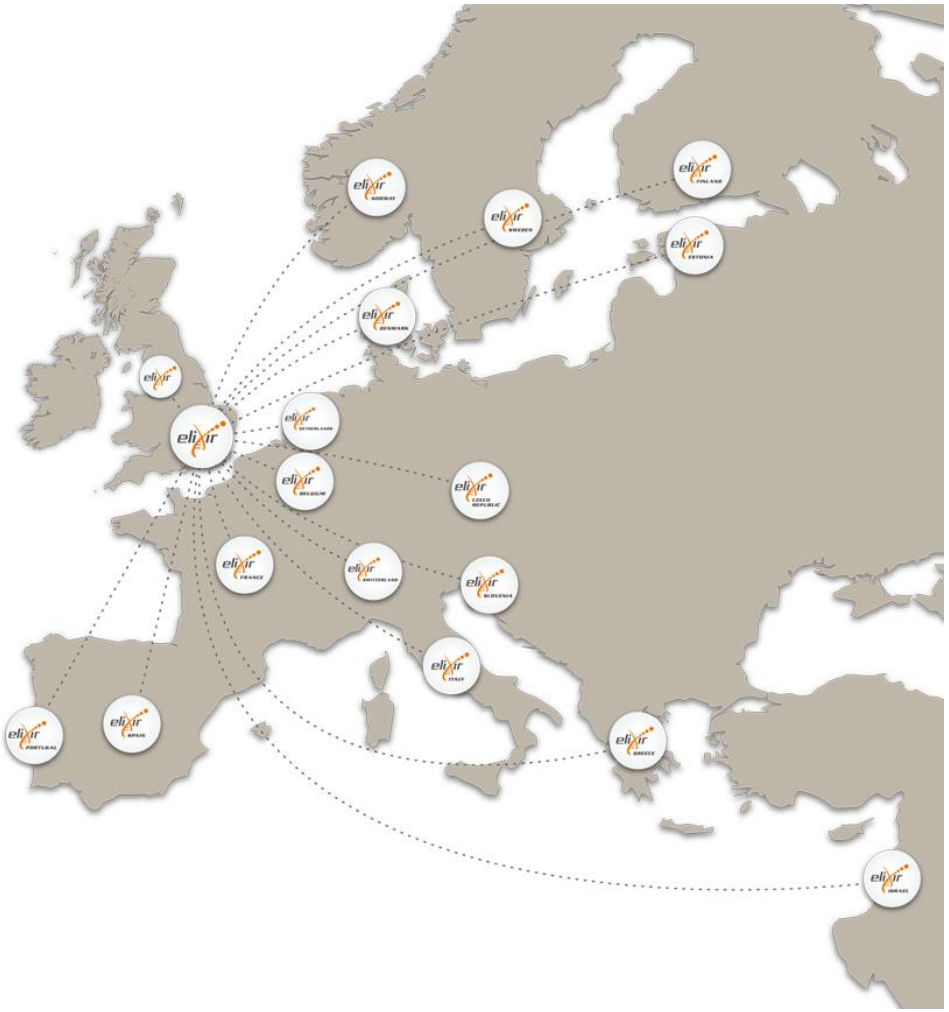
  Access, Exchange & Compute on sensitive data

- **Standards**

  Integration and interoperability of data and services.

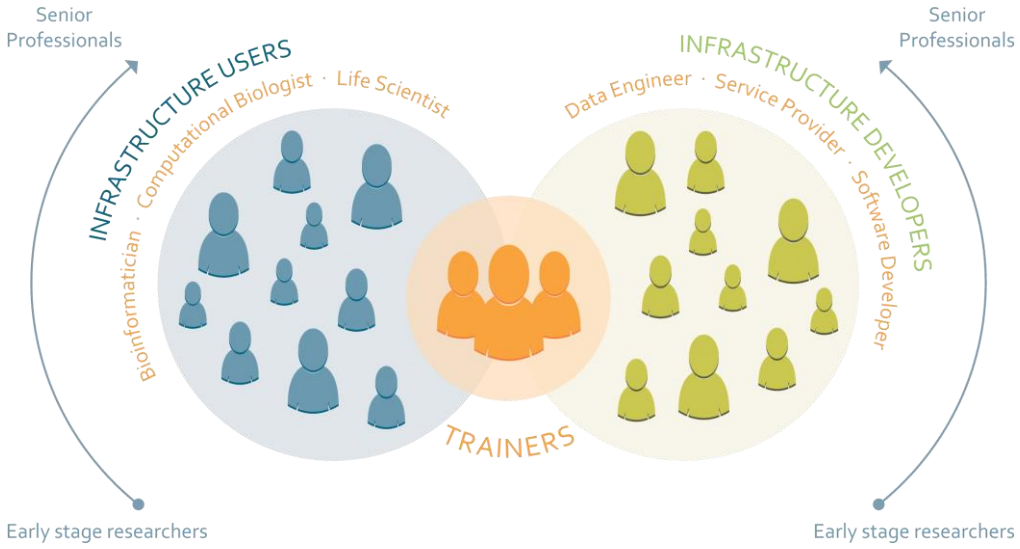
- **Training**

  Professional skills for managing and exploiting data



ELIXIR Training Strategy

- Facilitate accessibility to Europe's bioinformatics resources by up-skilling researchers who can more effectively exploit the data, tools, standards and compute services provided by ELIXIR
- Support and train users through e-learning, face-to-face courses and programs held across Europe
- Develop a coordinated pan-European training program of high quality and impact
- Partnership with global efforts such as GOBLET

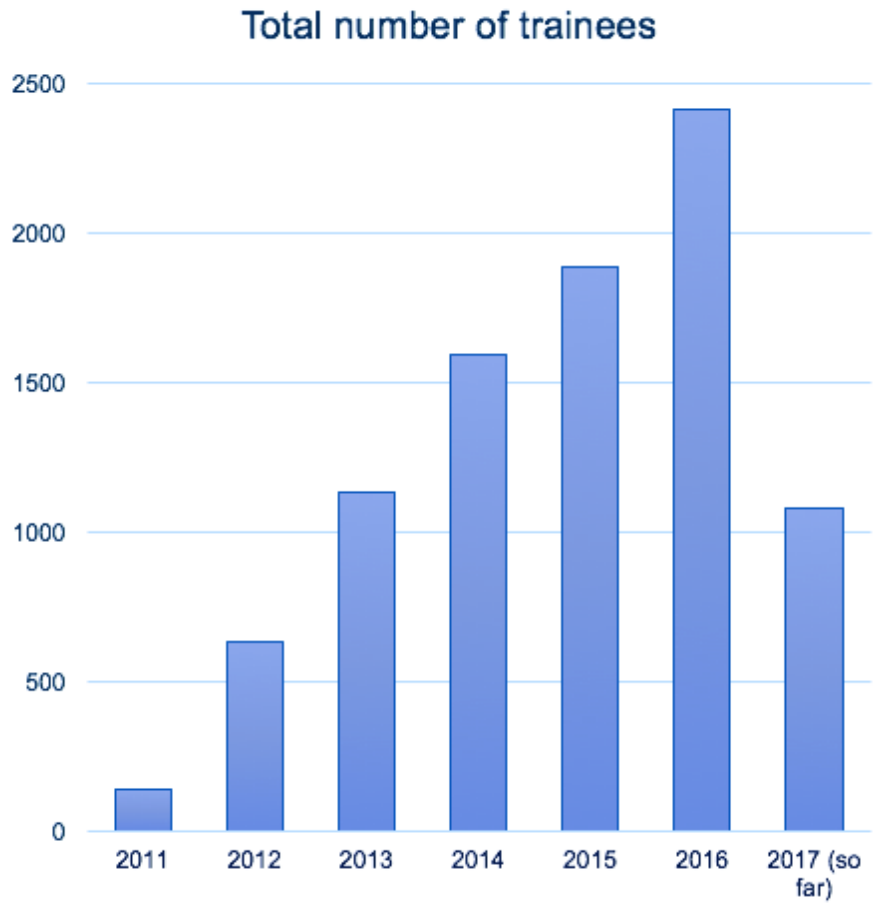


GOBLET - http://mygoblet.org/

- Global Organisation for Bioinformatics Learning, Education & Training

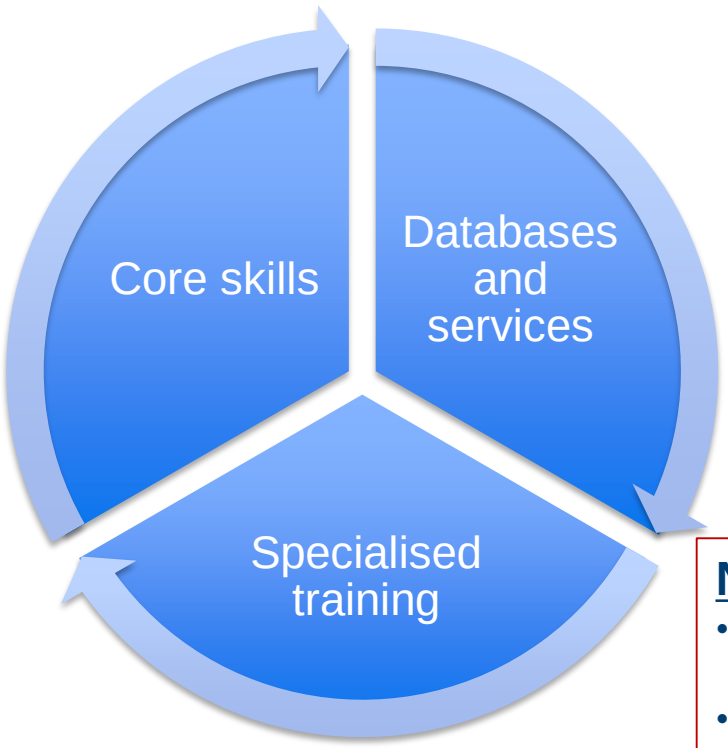
- Provide a global, sustainable support and networking structure for bioinformatics trainers and trainees, including (i) a training portal for sharing materials, tools, tips and techniques; (ii) guidelines and best practice documents; (iii) facilities to help “Train the Trainers”; and (iv) offering different learning pathways for different types of learner

Bioinformatics postgraduate training @ UoC



Bioinformatics postgraduate training @ UoC

- Introduction to:
  - ✓ MATLAB,
  - ✓ PERL,
  - ✓ Python and
  - ✓ R
- Data Carpentry
- Introductory statistics and experimental design for genomics
- Software Carpentry
- Statistical analysis using R



- Interpreting the clinical genome with Decipher
- Biological data analysis using Intermine
- EMBL-EBI courses:
  - ✓ GWAS catalog
  - ✓ Interactions & pathways – IntAct
  - ✓ Interactions & pathways – Reactome
  - ✓ Network analysis - Cytoscape and PSICQUIC
  - ✓ Ensembl API workshop

- Analysis of DNA methylation using sequencing
- Image analysis for biologists
- Introduction to Galaxy: RNA-seq and ChIP-seq data analysis
- Introduction to RNA-seq and ChIP-seq data analysis
- Molecular phylogenetics

New courses for 2017:

- Analysis of RNA-seq data with Bioconductor
- An introduction to Machine Learning
- Python: Functions and Modules; Best Practices
- Introduction to Machine Learning with R

What do we aim for?

- Course content. Courses consist of a well balance mixture of lectures and hands-on sessions
- Software choice. Focus on the use of open source, stable, actively developed and well-maintained software tools (i.e. Bioconductor, Galaxy,…)

- Objectives. Trainees should learn:
  - ✓ how to interpret biological data;
  - ✓ what a specific data analysis pipeline entails; and
  - ✓ how to critically evaluate the data analysis tools available.

- Objectives. We want to enable you to establish a partnership with your statistician and/or bioinformatician collaborators, based on mutual understanding

Bacterial Genome Assembly & Annotation in Galaxy– June 8

- Cover the basics of *de novo* genome assembly using small genome example
- Focusing on project planning, selecting fragment sizes, initial assembly of reads and then assembling those contigs
- Session includes an introduction to the Galaxy Platform
- Will give participants an understanding of how bacterial genome assembly works, how to run such analysis and be able to annotate a small genome
- Aims:

1. Introduction to Galaxy
2. Brief introduction to next generation sequencing data
3. Genome assembly and its parameters
4. Automatic genome annotation pipeline

Data Carpentry workshop – Sep 11/12

- Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research
- Focus is on the introductory computational skills needed for data management and analysis in all domains of research
- Target audience is learners who have little to no prior computational experience
- Topics:

1. Data organization in spreadsheets
2. Data cleaning with OpenRefine
3. Introduction to R
4. Data analysis and visualization in R
5. SQL for data management


RESOURCES
FUNCTIONS
CTR
BIOLOGY
BEST
BASIC
USING
EXPLORING
CHIP-SEQ
STATISTICS
STRUCTURE
PLACENTAL
TRAIN-THE-TRAINER
RNA-SEQ
PROBLEMS
CRUK
GATK
SCIENTIFIC
DATA
EXCELERATE
DISEASE
ANNOTATION
LIFE-SCIENCES
DNA
CELL
GALAXY
PRACTICES
APPLICATION
PROTEIN
CLINICAL
SOLVING
SINGLE
FREE
CELLPROFILER
STATISTICAL
IMAGE
DESIGNING
PACKAGE

Sources of training materials/information

- ELIXIR training portal: https://tess.elixir-europe.org/
- GOBLET training portal: http://mygoblet.org/training-portal
- Data Carpentry/Software Carpentry: http://www.datacarpentry.org/lessons/ and http://software-carpentry.org/lessons/
- See individual course pages on: http://bioinfotraining.bio.cam.ac.uk/
- Online resources: EMBL-EBI Train online, EdX (Data for life sciences from Harvard-Irizarry), Coursera, etc.
- Training course catalogue: https://www.on-course.eu/
- And many more…….