ELIXIR – the European infrastructure for life-science information

Célia Miguel
Trees4Future, Brussels, 5th April

www.elixir-europe.org
ELIXIR connects national bioinformatics centres and EMBL-EBI into a sustainable European infrastructure for biological research data.

ELIXIR underpins life science research across academia and industry.
An international distributed infrastructure for biological data

**Technical platforms**
- Data
- Standards
- Tools
- Compute
- Training
- Industry

**Use cases**
- Marine metagenomics
- Crop and forest plants
- Human data
- Rare diseases
A distributed infrastructure to scale with the challenge

**ELIXIR** data infrastructure for Europe’s life science research sector

**ELIXIR** Nodes build local bioinformatics capacity throughout Europe

**ELIXIR** Nodes build on national strengths and priorities
Members

Belgium  Czech Republic  Denmark

EMBL  Estonia  Finland

France  Israel  Italy

Netherlands  Norway  Portugal

Slovenia  Spain  Sweden

Switzerland  United Kingdom

Observers

Greece  Ireland

- 17 Members
- 2 Observers

Collaboration Strategy agreed between ELIXIR and de.NBI – prelude to membership
ELIXIR Nodes

- Over 140 institutes involved in ELIXIR Nodes
- Collaboration Agreements to be signed between Hub and Nodes
- Nodes that are national networks are developing own internal Node agreements
A European and global priority

"ELIXIR is nominated because it is a world-leading infrastructure, vital for enabling the life sciences to derive maximum knowledge and understanding from biological, medical and environmental 'Big Data'."

- ELIXIR on the ESFRI roadmap
- ELIXIR one of three prioritised Research Infrastructures
- ELIXIR defined by G7 as infrastructure with global reach

Group of Senior Officials on Global Research Infrastructures Progress Report 2015, October 2015
ELIXIR Services

Data deposition:
ENA, EGA, PDBe, EuropePMC, ...

Discovery tools:
Bio.tools

Data management:
Genome annotation
Data management plans

Value-added knowledge-bases:
UniProt, Ensembl, OrphaNet, ...

Compute:
Secure data transfer, cloud computing, AAI

Interoperability:
FAIR, Bring Your Own Data, Ontologies

Industry:
Innovation and SME programme
Bespoke collaborations

Training:
TeSS, Data Carpentry, eLearning
ELIXIR-EXCELERATE

- Develop and connect resources and services (from Nodes)
- Build bioinformatics capacity across Europe (Capacity Building and Training)
- Platforms - Data; Tools; Compute; Interoperability; Training
- Use Cases - Rare diseases; Human data; Plants; Marine metagenomics
- EUR 19 million; 41 partners
EXCELERATE Structure

Build ELIXIR platforms

WP1: Tools interoperability & registry
WP2: Community benchmarking
WP3: Data resources
WP4: Compute, Data access and exchange services
WP5: Interoperability backbone
WP11: Training

Use ELIXIR Services

WP6: Marine
WP7: Plants
WP8: Rare diseases
WP9: Human data

Embed and sustain ELIXIR
Integrating Genomic and Phenotypic Data for Crop and Forest Plants

- Population growth and climate change: major challenges to food security, forest management,…
- Traditional routes to crop improvement: too slow to keep up with demand.
- Plant genomes: valuable tool to understand important traits such as tolerance to drought, salt,…
Integrating Genomic and Phenotypic Data for Crop and Forest Plants

- phenotypic variation in most crop and forest plants is enormous, and phenotype characterisation data is
  
  (i) often inaccessible
  
  (ii) diverse and non-standard
  
  (iii) lacks any route of unified access

technical infrastructure and associated social practices needed to define an open model for publishing and sharing of plant genotype-phenotype data.
Use Case – Integrating Genomic and Phenotypic Data for Crop and Forest Plants

• Make data interoperable by the use of controlled vocabularies for annotating plant phenotypic data
  ➢ critical to enabling diverse datasets to be interrogated and compared

• Annotate a variety of data types and submit key exemplar datasets (from at least 3 plant species) to relevant public archives

Paul Kersey, EBI & Celia Miguel, PT
Use Case – Integrating Genomic and Phenotypic Data for Crop and Forest Plants

- Development and implementation of agreed public APIs for access to data and exposure via public computational infrastructures
- Engage industry in defining priorities in genotype/phenotype annotations
- Specific training for end users

Paul Kersey, EBI & Celia Miguel, PT
Workshop PhenoHarmonis 9-13 May, 2016 – Montpellier, France

Semantics for Harmonization and Integration of Phenotypic and Agronomic Data
9-13 May 2016 - Montpellier, France

Contents
1 Questions to be addressed
2 Structure of the workshop
3 Sponsors
Scientific collaborations – ELIXIR in EU projects

- **CORBEL (2015-2018)** – Harmonising access to resources and services of 11 biomedical research infrastructures
- **EMBRIC (2015-2018)** - connects marine biotechnology initiatives that focus on science, industry and regional growth.
- **EMTRAIN (2014-2016)** – to establish a pan-European platform for education and training for biomedical researchers
- **AARC (2015-2017)** – to develop an integrated cross-discipline AAI framework, built on production and existing federated access services.
- **ENVRIPlus (2015-2019)** - to create an interdisciplinary and interoperable cluster of Environmental Research Infrastructures across Europe.
ELIXIR tools and data registry [http://bio.tools](http://bio.tools)

Community curated registry with plant-specific databases and analysis tools

<table>
<thead>
<tr>
<th>PlantsDB</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://pgdb.helmholtz-muenchen.de/plants/index.jsp">http://pgdb.helmholtz-muenchen.de/plants/index.jsp</a></td>
</tr>
<tr>
<td>PlantsDB currently holds structured and interconnected genome and gene data for 12 major plant species ranging from model organisms like Arabidopsis to important crops, like wheat. Additionally it provides a platform for integrative and comparative plant genome research for which specialized online tools such as Ensembl, a comparative map viewer, or an RNAseqExpression Browser. Further useful features are the bulk data FTP download possibilities and PGDB Reloads a large plant transposon collection connected to REcat a detailed repeat classification scheme.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LAILAPS Search Engine (transPLANT instance)</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://webapps.plk-gatersleben.de/lailaps/app">http://webapps.plk-gatersleben.de/lailaps/app</a></td>
</tr>
<tr>
<td>LAILAPS is a search engine which provides an extraction of plant genomic data for forward genetic research and...</td>
</tr>
</tbody>
</table>
Examples of ELIXIR Nodes

www.elixir-europe.org
• BioData.pt being created
• Led by INESC-ID
• National funding award made
• Focus on woody plants
National strengths and priorities

Standards
Ontologies
Data Resources
Data Curation
Data Sharing
EMBL- EBI

- ‘European Node’
- EMBL-ELIXIR Work Programme agreed, ie, Ensembl, EGA, ENA, PDBe, EuropePMC
- Lead EXCELERATE Use Cases on Rare Diseases, Human Data, Plant Sciences, Marine..
- Involved in several ELIXIR Pilot Actions
ELIXIR Norway

- National network exists for number of years
- Focus on marine informatics and clinical data
- Leads EXCELELRATE Use Case on Marine Metagenomics
ELIXIR France

- New national infrastructure – IFB
- CNRS, INRA, INSERM, Pasteur...
- New investment from national roadmap – EUR 19 million
- Marine, microbes, plant sciences, health...
ELIXIR Italy

- Represented through Joint Research Unit
- New national network with many partners
- Includes e-Infrastructure partners ie CINECA
- Submitted Node proposal during Observer status
ELIXIR UK

- Focus on training
- Data Carpentry and Software Carpentry
- TeSS portal
- Supported by grant through Research Councils
- Node currently undergoing expansion
Further information
ELIXIR Gateway on F1000Research

• Online platform to capture ELIXIR’s publication output, research articles transparently peer-reviewed through F1000R

• ELIXIR Editorial board:
  • Chair: Niklas Blomberg, Members - ELIXIR Heads of Nodes: Arlindo Oliveira, Barend Mons, Inge Jonassen, Bengt Persson

• Latest articles:
  • Moreews F et al. BioShaDock: a community driven bioinformatics shared Docker-based tools registry
  • Eijssen et al. The Dutch Techcentre for Life Sciences: Enabling data-intensive life science research in the Netherlands
  • Bousfield et al. Patterns of database citation in articles and patents indicate long-term scientific and industry value of biological data resources (Pending review)

http://f1000research.com/channels/elixir
Get involved – keep in touch

**Informed: ELIXIR eNewsletter**

Dear reader,

Welcome to the spring 2015 edition of *Informed*, the ELIXIR eNewsletter.

ELIXIR's mission is to build a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine and the environment, the bio-industries and society. *Informed* provides you with the latest news and updates on developments within the ELIXIR initiative. I hope you enjoy reading.

Dr Niklas Blomberg
ELIXIR Director

Sign up at:
https://www.elixir-europe.org/news
Thank you