ELIXIR - CORBEL and contributions of the EU health related infrastructures to Personalised Medicine in Europe

Jerry Lanfear, ELIXIR CTO

INFRAFRONTIER Stakeholder Meeting, November 2017

www.elixir-europe.org
ELIXIR, a distributed infrastructure for biological data

Note: ELIXIR Hub uses EMBL’s legal model => ELIXIR categorised as international treaty organisation
ELIXIR funding

Members/Nodes funded by national agencies

Implementation studies

Hub funded jointly by ELIXIR member states

Hub & Nodes - Joint applications to EU funding
ELIXIR in numbers

- 21 Members and 1 Observer
- ~180 institutes involved
- 600+ staff
- 16 Core Data Resources
- 23 Implementation Studies ongoing or soon to start
- 17 papers in ELIXIR F1000 channel
- 264 live events in TeSS
- 350 companies attended Innovation and SME programme
**ELIXIR: European infrastructure for biological information**

Data infrastructure for Europe’s life-science research:

<table>
<thead>
<tr>
<th>Platforms</th>
<th>Use cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Marine metagenomics</td>
</tr>
<tr>
<td>Interoperability</td>
<td>Crop and forest plants</td>
</tr>
<tr>
<td>Tools</td>
<td>Human data</td>
</tr>
<tr>
<td>Compute</td>
<td>Rare diseases</td>
</tr>
<tr>
<td>Training</td>
<td></td>
</tr>
</tbody>
</table>

[@ELIXIREurope](https://twitter.com/ELIXIREurope)  [www.elixir-europe.org](http://www.elixir-europe.org)
Vision

To facilitate discoverability, access, sharing and analysis of genomics data, including rare disease, linked to other data types, at scale (4-5M participants)

To demonstrate how use of infrastructure can impact translation of genomics research into medicine
Infrastructure for Human Genomics and Translational data

Understand disease

Patient Stratification and PM

Develop and test novel pharmacological hypotheses

Unleash the possibilities for genomics and health
Mission

- To construct and operate a sustainable infrastructure for Human Genomics and Translational data in Europe to support life science research and its translation to medicine
“Whenever possible, biological research data should be submitted to the recommended community deposition databases”

<table>
<thead>
<tr>
<th>Deposition Database</th>
<th>Data type</th>
<th>International collaboration framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArrayExpress</td>
<td>Functional genomics data. Stores data from high-throughput functional genomics experiments.</td>
<td></td>
</tr>
<tr>
<td>BioModels</td>
<td>Computational models of biological processes.</td>
<td></td>
</tr>
<tr>
<td>EGA</td>
<td>Personally identifiable genetic and phenotypic data resulting from biomedical research projects.</td>
<td>European Bioinformatics Institute and the Centre for Genomic Regulation</td>
</tr>
<tr>
<td>ENA</td>
<td>Nucleotide sequence information, covering raw sequencing data, contextual data, sequence assembly information and functional and taxonomic annotation.</td>
<td>International Nucleotide Sequence Database Collaboration</td>
</tr>
<tr>
<td>IntAct</td>
<td>IntAct provides a freely available, open source database system and analysis tools for molecular interaction data.</td>
<td>The International Molecular Exchange Consortium</td>
</tr>
<tr>
<td>MetaboLights</td>
<td>Metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from metabolic experiments.</td>
<td></td>
</tr>
<tr>
<td>PDBe</td>
<td>Biological macromolecular structures.</td>
<td></td>
</tr>
<tr>
<td>PRIDE</td>
<td>Mass spectrometry-based proteomics data and mRNA and protein expression information (with fold change and/or p-values) and the submission of new mass spectrometry data.</td>
<td></td>
</tr>
</tbody>
</table>

- The ELIXIR Deposition Databases meet the technical quality and governance criteria expected of ELIXIR Core Data Resources
- Agreed collectively by 21 Heads of Nodes
- International collaborative effort

https://elixir-europe.org/platforms/data/elixir-deposition-databases
ELIXIR Core Data Resources

- Set of European data resources of fundamental importance to the wider life-science community and the long-term preservation of biological data.
- Agreed collectively by 21 Heads of Nodes
- International collaborative effort

https://elixir-europe.org/platforms/data/core-data-resources

<table>
<thead>
<tr>
<th>ELIXIR Core Data Resource list</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core Data Resource</td>
</tr>
<tr>
<td>--------------------</td>
</tr>
<tr>
<td>ArrayExpress</td>
</tr>
<tr>
<td>CATH</td>
</tr>
<tr>
<td>CHEBI</td>
</tr>
<tr>
<td>CHEMBL</td>
</tr>
<tr>
<td>EGA</td>
</tr>
<tr>
<td>ENA</td>
</tr>
<tr>
<td>Ensembl</td>
</tr>
<tr>
<td>Ensembl Genomes</td>
</tr>
<tr>
<td>Europe PMC</td>
</tr>
<tr>
<td>Human Protein Atlas</td>
</tr>
<tr>
<td>The IMEx Consortium:</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>InterPro</td>
</tr>
</tbody>
</table>
European Genome-phenome Archive (EGA)

Controlled access for human genomic data

Data submitters: >600
(Universities, research institutes, consortia, pharma, ...)

Data requesters:
> 7,100 (worldwide)
Controlled access:
Data Access Committees

www.ega-archive.org
Local EGA

- Sensitive data are stored locally
- EGA provides the software platform
- First prototype in 2017 with national partners
Beacons: Public data discovery web-service

Beacon X: Yes
Beacon Y: No
Beacon Z: No...

Do you have information about the allele “C at position 32,936,732 on chromosome 13?”

Yes / No
(+optional metadata about the allele)

beacon-network.org/
ELIXIR Interoperability

Interoperability services and practices to support FAIR data and interoperability activities

Use case focused
Data providers
Data integrators

With international initiatives, from community grassroots to government programmes.
• Schema.org = **Structured data markup for web pages**

• Bioschema.org = schema.org for biological information (dataset, sample, standard, protein, training material ...)
  • Specification on top of schema.org
  • Layer of constrains + documentation + extensions

• Minimum properties for finding data
ELIXIR Position Paper on FAIR data management in the life sciences

1. Open sharing of research data is a core principle
2. Data Management is crucial to science
3. Data should be submitted to deposition databases
4. All data submitted to Open Data archives should align with community-defined standards
5. ELIXIR Nodes implement FAIR for their respective nations
6. Professional skills, adequate resources and appropriate funding are needed for Data Management and infrastructure

Practical demonstration: Data storage and transfer, coupled to security

- **ELIXIR Webinar:** Transfer of large volumes of confidential human data, while maintaining appropriate access rights
- 13 November 2017
- Slides and recording available at: [https://www.elixir-europe.org/events/elixir-webinar-transfer-large-volume-data](https://www.elixir-europe.org/events/elixir-webinar-transfer-large-volume-data)
ELIXIR Authorisation and Authentication Infrastructure

Enables life science researchers to use their institutional IDs to access services and data:

- **Reduced bureaucracy and costs**
- **Improved vetting:** federated identities provide greater confidence to the service and data providers
- **Regular updates:** as researchers join and leave institutions, their affiliation information is maintained regularly
- **Improved access to usage metrics:** consistent use of accounts allows service providers to better analyse the use of their services
- **Applicable to other research infrastructures (CORBEL)**
Thirteen ESFRI Research Infrastructures in Bio-Medical Sciences to **harmonise access and services** for complex research projects involving more than one research infrastructure.

- biobanking & biomolecular resources
- curated databases
- highly pathogenic microorganisms
- functional genomics
- microorganisms
- translational research
- marine model organisms
- screening & medicinal chemistry
- structural biology
- clinical trials
- plant phenotyping
- biological/medical imaging
- systems biology
BACKGROUND

- 4 year project: 2015-2019
- 37 partners in 13 BMS RIs
- budget: €14.8 million
- co-coordinated by ELIXIR and BBMRI-ERIC
**Use Case 1**
Genotype-to-Phenotype analysis based on models and experimental data

**Use Case 2**
Predictive systems pharmacology for safer drugs and chemical products
ELIXIR: European infrastructure for biological information

Data infrastructure for Europe’s life-science research:

Platforms | Use cases
---|---
Data | Marine metagenomics
Interoperability | Crop and forest plants
Tools | Human data
Compute | Rare diseases
Training

@ELIXIREurope  www.elixir-europe.org