FAIR data, tools and workflows: a user perspective

Frederik Coppens

NETTAB 2018 - Building a FAIR Bioinformatics environment

www.elixir-europe.org
Outline

ELIXIR: European infrastructure for biological information

Open Science, Open Data & Research Data Management

ELIXIR Services

Infrastructure for Open Science
ELIXIR: European infrastructure for biological information
European infrastructure for biological information

Domain specific Services
ESFRIs

Storage, Compute, Network
e-Infrastructures
Data infrastructure for Europe’s life-science research
European-wide effort
Open Science, Open Data & Research Data Management
The theory

HOW IT WORKS

- Get funding
- Write DMP
- Gather data during research
- Finish research

Publish results

Choose repository and deposit data

Inform OpenAIRE the EU-funded Open Access portal
Issues with the theory for researchers

- What should be in a Data Management Plan?
- Gathering data? What to gather?
- What is metadata? Never heard of...
- Which repository to choose?
- How to deposit (large amounts of) data?
- What is OpenAire?
In practice
ELIXIR Services

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

Data management:
Genome annotation
Data management plans

Added value data:
UniProt, Ensembl, OrphaNet, ...

Data Interoperability:
BioSharing, identifiers.org and OLS

Compute:
Secure data transfer, cloud computing, AAI

Bioinformatics tools:
Bio.tools

Industry:
Innovation and SME programme
Bespoke collaborations

Training:
TeSS, Data Carpentry, eLearning
Data infrastructure for Europe’s life-science research
FAIR Data Position Paper

1. **Open sharing of research data is a core principle** for publicly-funded research and ELIXIR encourages all funders to adopt Open Data mandates.

2. **Data Management is crucial** part of good scientific practice and research excellence.

3. Whenever possible, biological research data should be submitted to the recommended community **deposition databases**.

4. All data submitted to Open Data archives must be **annotated** in accordance with **community-defined standards**.

5. **ELIXIR Nodes** are the national implementation of a harmonised FAIR Data Management programme for the life sciences.

6. FAIR data management requires **professional skills** and adequate resources.

7. Good research data management requires appropriate **funding for data infrastructures**.

Data Platform

Deposition Databases for Biomolecular Data

- ArrayExpress: Functional genomics data
- BioModels: Computational models of biological processes
- EGA: Genetic and phenotypic data from biomedical research projects
- ENA: Nucleotide sequence information
- IntAct: Molecular interaction data
- MetaboLights: Experimental metabolomics data
- PDBe: Biological macromolecular structures
- PRIDE: Mass spectrometry-based proteomics data

https://www.elixir-europe.org/platforms/data/elixir-deposition-databases
Data Platform

ELIXIR Core Data Resources

Fundamental importance to the wider life-science community

Long-term preservation of biological information

Ensure sustainability

Examples:

- European Nucleotide Archive (ENA)
- Europe PubMed Central
- Human Protein Atlas

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

Drive implementation of the Beacon technology within ELIXIR nodes

ELIXIR Authentication and Authorisation Infrastructure

<table>
<thead>
<tr>
<th>Type</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>Accessible to internet users</td>
</tr>
<tr>
<td>Registered</td>
<td>Accessible to bona fide researchers</td>
</tr>
<tr>
<td>Controlled</td>
<td>Authorized, signed agreement needed</td>
</tr>
</tbody>
</table>
OmicsDI

- Omics Discovery Index
- integrated and open source platform
- facilitating the access and dissemination of omics datasets
  - proteomics
  - genomics
  - transcriptomics
  - metabolomics

www.omicsdi.org
OmicsDI

- 452.6K Datasets
- 65.2K Species
- 8.5K Multiomics
- 1.6K Tissues
- 308.3K Genomics
- 327 Users
- 124.3K Transcriptomics
- 19 Repositories
- 2.6K Diseases
Interoperability Platform

BioSchemas.org

- **Schema.org** semantic markup of webpages
- collection of specifications that provide guidelines for markup within life sciences
- Generic service as part of the European Open Science Cloud
<table>
<thead>
<tr>
<th>Recipe Title</th>
<th>Source</th>
<th>Rating</th>
<th>Calories</th>
<th>Nutritional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple Pie by Grandma Ople Recipe - Allrecipes.com</td>
<td>allrecipes.com/recipe/126832/apple-pie-by-grandma-ople/</td>
<td>★★★★★</td>
<td>512 cal</td>
<td>Total Fat: 11 g, Saturated fat: 3.6 g, Polyunsaturated fat: 2.2 g, Monounsaturated fat: 4.4 g, Cholesterol: 0 mg, Sodium: 266 mg</td>
</tr>
<tr>
<td>Perfect Apple Recipe - Pillsbury</td>
<td><a href="https://www.pillsbury.com/...apple-pie/162b605-0af4f-441e-ad93-bbd0">https://www.pillsbury.com/...apple-pie/162b605-0af4f-441e-ad93-bbd0</a>...</td>
<td>★★★★★</td>
<td>230 cal</td>
<td></td>
</tr>
<tr>
<td>Scrumptious Apple Pie Recipe - BettyCrocker.com</td>
<td><a href="https://www.bettycrocker.com/...apple-pie/9a4acc6-85aa-4128-b0b0-1a">https://www.bettycrocker.com/...apple-pie/9a4acc6-85aa-4128-b0b0-1a</a>...</td>
<td>★★★★☆</td>
<td>480 cal</td>
<td></td>
</tr>
<tr>
<td>Apple Pie Recipe</td>
<td>Food Network Kitchen</td>
<td>Food Network</td>
<td><a href="http://www.foodnetwork.com/recipes/food-network.../apple-pie-recipe-201142">www.foodnetwork.com/recipes/food-network.../apple-pie-recipe-201142</a>...</td>
<td>★★★★★</td>
</tr>
</tbody>
</table>
Training eSupport System

- Disseminate, discover & package training resources, training materials and events
- Aggregating information from ELIXIR nodes and 3rd-party content providers

https://tess.elixir-uk.org
FAIRsharing.org
(formerly BioSharing.org)

FAIRsharing is here! From our first incarnation, BioSharing.org, which focussed on the life sciences, we are growing into FAIRsharing.org, to serve users across all disciplines and support Findable, Accessible, Interoperable and Reusable (FAIR) data.

A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.

Find

Recommendations

Standards and/or databases recommended by journal or funder data policies.

Discover

Collections

Standards and/or databases grouped by domain, species or organization.

Learn

Educational

About standards, their use in databases and policies, and how we can help you.
In practice
Tools Platform

*improve the discovery, quality and sustainability of software resources*

Tools and services registry

Scientific benchmark and technical monitoring

Software deployment

Workflows and workbenches

Software best practices

Jiménez et al. *Four simple recommendations to encourage best practices in research software. F1000Research* 2017, 6:876

Tools interoperability (e.g. Common Workflow Language)
Welcome to bio.tools! Please check out our Documentation. There is a lot of useful information there.

This site uses cookies. By continuing to browse the site you are agreeing to our use of cookies. Find out more here.

WebGestalt

WebGestalt (WEB-based GEne Set Analysis Toolkit) is a system facilitating the analysis of sets of genes. Gene sets can be compared using set operations (intersection, union, etc.), different annotations can be selected and retrieved for the set, and sets can be visualized and organized by a user-selected method (Gene Ontology, chromosomal distribution, etc.). WebGestalt can also perform a statistical analysis to suggest areas of interest with respect to the set of genes selected. Pathway maps.

- Molecular interactions, pathways and networks
- Model organisms
- Mapping
- Transcriptomics
- Gene expression
- Web application

FFGED

The filamentous fungal gene expression database (FFGED) is a user-friendly management of gene expression data, which are assorted into experimental metadata, experimental design, raw data, normalized details, and analysis results. FFGED functions as a collective and collaborative platform, by connecting each experiment with similar related experiments made public by other users, and correlating diverse gene expression levels under multiple experimental designs within different experiments.

- Gene expression
- Experimental design and studies
- Gene transcripts
- Microarray experiment
- Transcription factors and regulatory sites
- Web application
Workflows
Galaxy

https://usegalaxy.org
ELIXIR Galaxy Community

Co-chairing with Björn Grüning (de.NBI, ELIXIR Germany) and Gildas Le Corguillé (FB, ELIXIR France)
Software deployment

https://biocontainers.pro/
In practice
Data Management WG

Data Stewardship Wizard

- Build a network of Data Stewards
- Smart questionnaire
- https://dsw.fairdata.solutions/
## My plans

The table below lists the plans that you have created, and any that have been shared with you by others. These can be edited, shared, exported or deleted at any time.

<table>
<thead>
<tr>
<th>Name</th>
<th>Owner</th>
<th>Shared?</th>
<th>Last edited</th>
<th>Select an action</th>
</tr>
</thead>
<tbody>
<tr>
<td>My project (FWO DMP)</td>
<td>Simon Stael</td>
<td>Yes (with 1 people)</td>
<td>24-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>My project (FWO DMP)</td>
<td>Amira M’Harmdi</td>
<td>Yes (with 2 people)</td>
<td>24-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>Frederik Coppens 2018</td>
<td>Me</td>
<td>Yes (with 1 people)</td>
<td>05-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>IRI 2018</td>
<td>Kim De Ruyck</td>
<td>Yes (with 1 people)</td>
<td>05-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>Unravelling temperature signalling in wheat</td>
<td>Ivo De Smet</td>
<td>Yes (with 1 people)</td>
<td>14-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>My project (FWO DMP)</td>
<td>Alexandra Baekelandt</td>
<td>Yes (with 1 people)</td>
<td>20-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>Inference of gene functions and regulatory annotations in dicotomes, a systems biology approach</td>
<td>Klaas Vandepoele</td>
<td>Yes (with 1 people)</td>
<td>08-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>My project (FWO DMP)</td>
<td>Bartel Vanholme</td>
<td>Yes (with 1 people)</td>
<td>13-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>Characterization of the plant Hippo pathway as an early auxin signaling module</td>
<td>Steffien Vanneste</td>
<td>Yes (with 1 people)</td>
<td>13-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>Characterisation of auxin-induced phosphorylation of auxin transporters in plant growth and development</td>
<td>Steffien Vanneste</td>
<td>Yes (with 1 people)</td>
<td>13-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>Identification of the major determinants that explain intolerance to gene duplication in the flowering plants</td>
<td>Eylem Aydogdu Lohaus</td>
<td>Yes (with 2 people)</td>
<td>13-03-2018</td>
<td>View Export</td>
</tr>
<tr>
<td>My project (FWO DMP)</td>
<td>Laurens Pauwels</td>
<td>Yes (with 1 neocia)</td>
<td>13-03-2018</td>
<td>View Export</td>
</tr>
</tbody>
</table>
Robotic phenotyping platforms @ VIB

- Designed and built in collaboration with machine building company - SMO

www.wiwam.com
Plant Phenotyping Data management

Phenotyping platforms  Image processing  Data analysis and visualization
Plant Phenotyping Data management
FAIRDOMHub

The FAIRDOMHub is built upon the SEEK software suite, which is an open source web platform for sharing scientific research assets, processes and outcomes. For more information about SEEK please visit http://seek4science.org

Please cite us in your publications if you use the FAIRDOMHub. For the citation to use please visit: https://fair-dom.org/cite

FAIRDOM is an initiative to develop a community, and establish an internationally sustained Data and Model Management service to the European Systems Biology community. FAIRDOM is a joint action of ERA-Net EraSysAPP and European Research Infrastructure ISBE

For more information about FAIRDOM please visit http://fair-dom.org

If you are interested in using FAIRDOMHub within your own funding programme, or have any other questions related to FAIRDOM and SEEK, including feature requests or how to get involved, please contact us at support@fair-dom.org

Get started

Who is using it?

Get involved
Infrastructure for Open Science
Infrastructure for Open Science

- Data: input, intermediate & results
- Tools & workflows
- All associated metadata
- Infrastructure to
  - Store
  - Analyse
  - Publish / share
- Data Management Platform
Infrastructure for Open Science
Data infrastructure for Europe’s life-science research

Life-science gateway for the European Open Science Cloud

Broadly applicable solutions for common challenges

FAIR at the Core
Challenge

Organize local data & metadata, compatible with daily research and publishing practices