EOSC-Life

Open, collaborative space for data-driven life science research

Nebras Gharbi, *Chief Data Officer*

Institut National de la Santé et de la Recherche Médicale

Data reuse in biomedical research (1/2)

- Isolated data production = untapped potential

- Open data (*with solutions to effectively use data, tools, and workflows*) enables new discoveries

- Maximizing research investment and accelerating scientific advancements

Data reuse in biomedical research (2/2)

- <u>Repository usage:</u>
  - The Gene Expression Omnibus (GEO) repository reported over **2 million** dataset downloads in 2020.
  - As of 2021, the European Nucleotide Archive was growing by **2 petabases per month**, indicating significant data sharing and potential for reuse.

- <u>Efficiency gains:</u>
  - The NIH Data Commons Pilot Phase Consortium reported that data reuse could reduce the time to initiate research projects by **up to 18 months**.

- <u>Discoveries:</u>
  - The Human Genome Project data, which is openly available, has been cited in over **5,200 patents**, demonstrating its impact on innovation.
  - A study of cancer genomics data reuse found that publicly available datasets led to **85 new cancer driver genes** being discovered.

- <u>Financial gains:</u>
  - A report by the European Commission estimated that the economic value of open data in the EU could reach **€334 billion by 2025**, with health being one of the key sectors.

EOSC-Life Overview

- 13 Biological and Medical ESFRI Research Infrastructures

- 23.7M€ budget, 4-year project (2019-2023)

- **Goal**: Enable data-driven life science research in EOSC
  - To improve access to a rich, diverse pool of life science data
  - To provide solutions that enable life scientists to effectively use data, tools, and workflows in the cloud
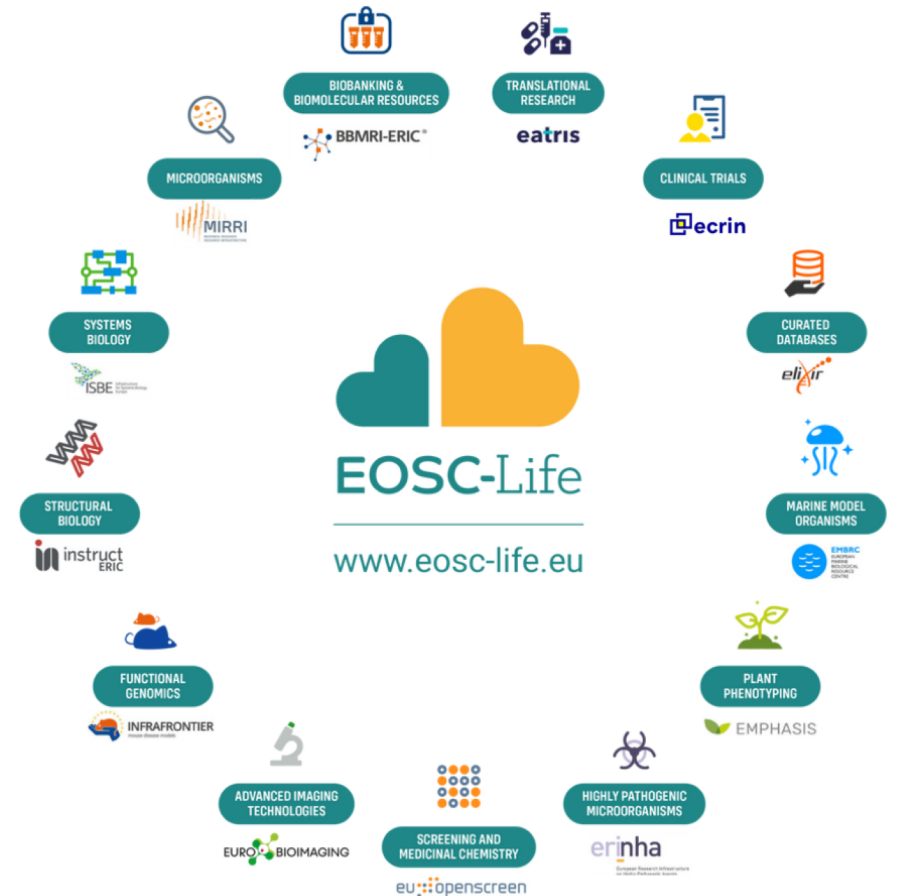
Key Achievements

- FAIR Data Resources

- Policies & Tools for Sensitive Data

- Ecosystem of Innovative Tools

- Connecting Scientists to EOSC (funding demonstrator projects, calls for projects, Training Open Calls)

FAIR Data Resources

- LS RI Data Catalogue: Over 2 million datasets from 13 RIs published in EOSC

- RO-Crate adopted as a standard for FAIR Digital Objects

- Provenance Information Standards (ISO 23494) developed

- COVID-19 Data Portal: Rapid data sharing platform for pandemic research

----

- FAIR-IMPACT project: Extending FAIR data practices across disciplines

- Data Management Expert Network established

Tools for Sensitive Data

- Sensitive Data Toolkit developed for ethical and legal compliance

- GDPR-compliant Federated EGA network established

- Secure data analysis environments created (e.g., Galaxy, tranSMART)

- ELIXIR AAI implemented for secure authentication across services

- Data Protection Impact Assessment tool developed
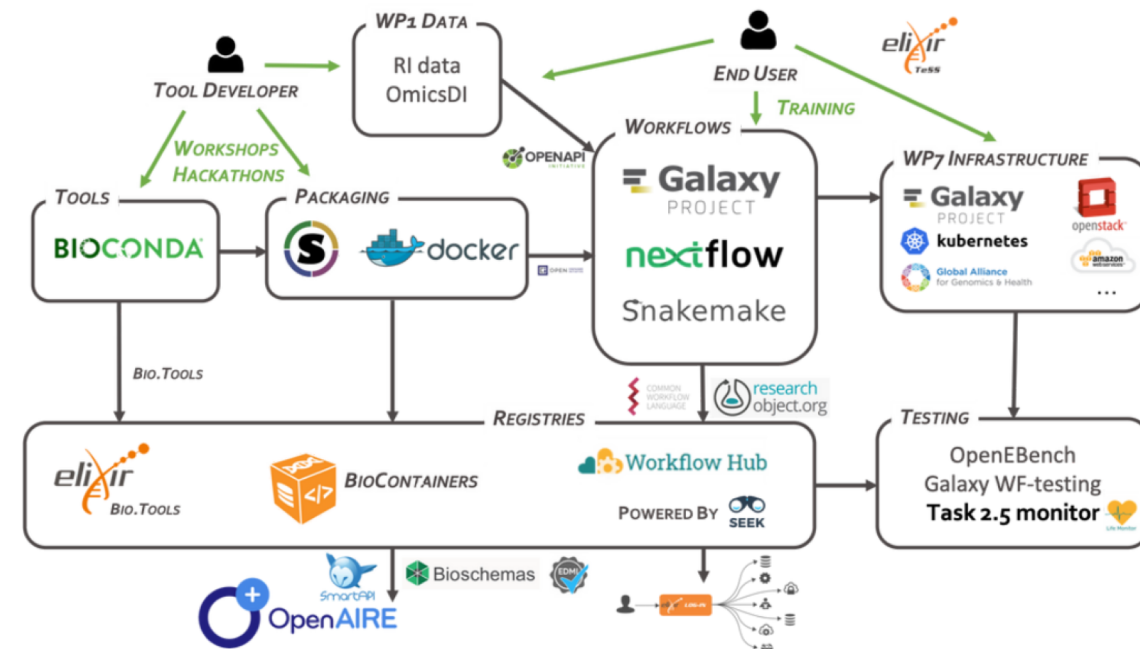
Ecosystem of Innovative Tools

- WorkflowHub.eu: Central registry for 1000+ computational workflows

- Bio.tools: Registry of 22,000+ life science tools and databases

- EOSC-Life Tools Collaboratory: 11 Open Call projects funded

- Workflow execution environments deployed (e.g., Galaxy, CWL)

- Common API implemented for improved tool discoverability

- Training programs developed: >4000 researchers trained

EOSC Life - Impacts

- Foundation for pandemic preparedness

- Support for cancer research (EOSC4Cancer)

- Basis for European Genomics Data Infrastructure

- Blueprint for data-driven research in EOSC

Resources