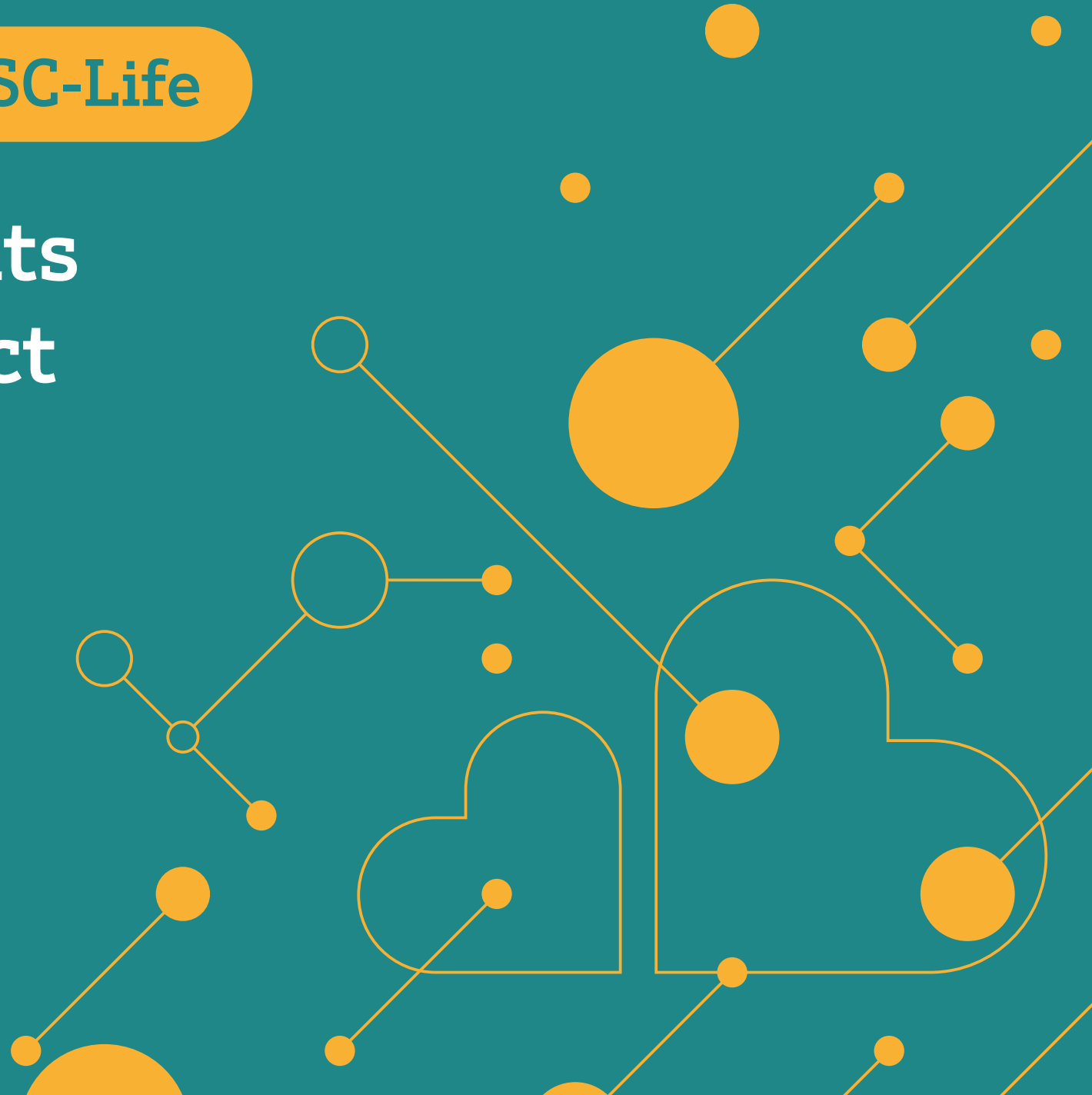## Populating EOSC-Life

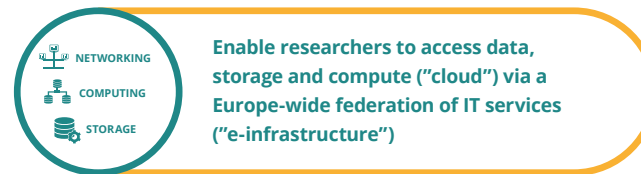# Achievements of the project thus far

EOSC-Life

EOSC-Life is creating an open, collaborative space for digital life science

EOSC-Life is a Horizon 2020-funded bringing together the 13 'ESFRI' Life Science Research Infrastructures (LS RIs) to create an open, digital and collaborative space for life science research.

The project co-creates and integrates the EOSC federated core, while simultaneously creating, adapting and adopting policies for Open Science.
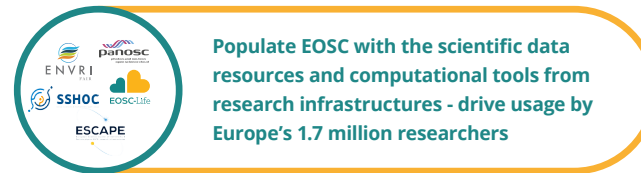
European Open Science Cloud

↓

NETWORKING
COMPUTING
STORAGE

Enable researchers to access data, storage and compute ("cloud") via a Europe-wide federation of IT services ("e-infrastructure")

E-Infrastructure consolidation

+

Drive the transition to Open Science (Open Data, Open Standards, Open Literature), to bring research benefits to European societies at large

Open Science

+

Populate EOSC with the scientific data resources and computational tools from research infrastructures - drive usage by Europe's 1.7 million researchers

Scientific **C**ommunities' content and users

All activities aim to fulfil our 4 project goals:

Establish EOSC-Life by publishing FAIR life science data resources in EOSC

Provide the policies, guidelines and processes for secure and ethical data reuse

Populate an ecosystem of innovative life science tools in EOSC

Enable data-driven research in Europe by connecting life scientists to EOSC via open calls for participation

Life science data is rich and diverse

The Life Science Research Infrastructures provide access to advanced instruments and research facilities that describe biology from single molecules to ecosystems and long-term population cohorts.

Connecting digital life science data will allow us to address major challenges facing humanity

These include the 5 EU missions in Horizon Europe:

- *Cancer*
- *Adaptation to climate change including societal transformation*
- *Healthy oceans, seas, coastal and inland waters*
- *Climate-neutral and smart cities*
- *Soil health and food*

EOSC-Life provides solutions so that life scientists can make use of data, tools and workflows in the cloud

Vast amounts of data are processed and analysed daily in the life sciences. EOSC-Life aims to make data, tools and analysis workflows more findable, accessible, interoperable and reusable (FAIR) through cloud deployment of these resources.

EOSC-Life
www.eosc-life.eu

MICROORGANISMS — MIRRI

BIOBANKING & BIOMOLECULAR RESOURCES — BBMRI-ERIC

TRANSLATIONAL RESEARCH — eatris

CLINICAL TRIALS — ECRIN

CURATED DATABASES — elixir

MARINE MODEL ORGANISMS — EMBRC

PLANT PHENOTYPING — EMPHASIS

HIGHLY PATHOGENIC MICROORGANISMS — erinha

SCREENING AND MEDICINAL CHEMISTRY — eu openscreen

ADVANCED IMAGING TECHNOLOGIES — EURO BIOIMAGING

FUNCTIONAL GENOMICS — INFRAFRONTIER

STRUCTURAL BIOLOGY — instruct ERIC

SYSTEMS BIOLOGY — ISBE

Structured into 14 work packages—including 2 new work packages in response to COVID-19—EOSC-Life will provide a unique opportunity for scientists to access and integrate an unprecedented richness of data resources. The use of common metadata specifications, cataloguing and indexing in data catalogues will make (LS RI) datasets widely accessible across disciplines.
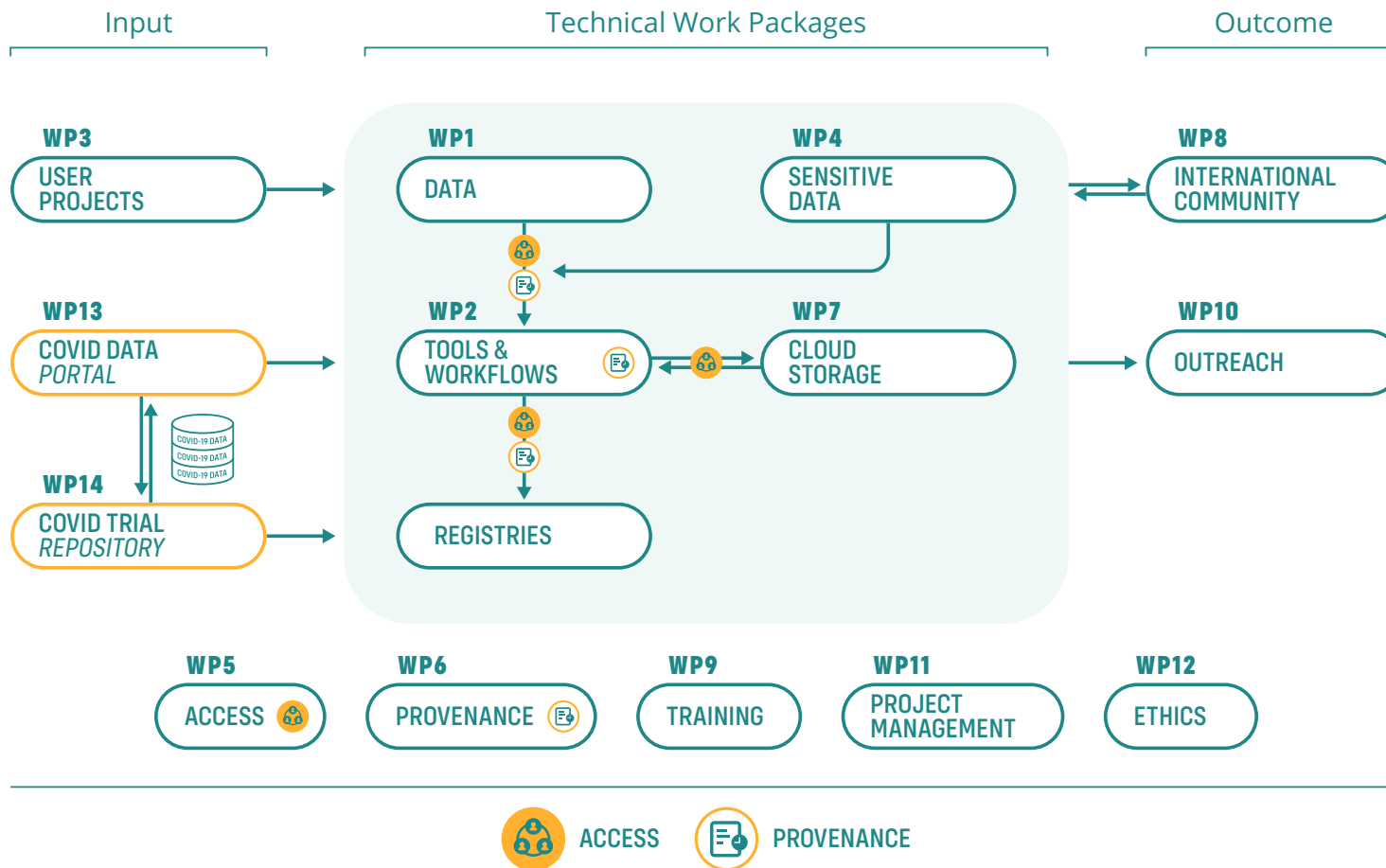
EOSC-Life's response to the COVID-19 pandemic

Data sharing and digital tools are critical for the response to the pandemic. EOSC-Life partners and solutions were well positioned to respond quickly to the needs of our community with efforts such as:

• *International COVID-19 biohackathon and cloud resources*
• *Frontloading of open tools and workflows (**WorkflowHub**, etc.)*
• *Building a repository to host COVID-19 clinical trial data*
• *Extension of the **COVID-19 Data Portal***

EOSC-Life

**Compliance with
EU regulations**

**COVID-19 clinical
trial data**

COVID-19 *Data Portal*

**Scientific
community**
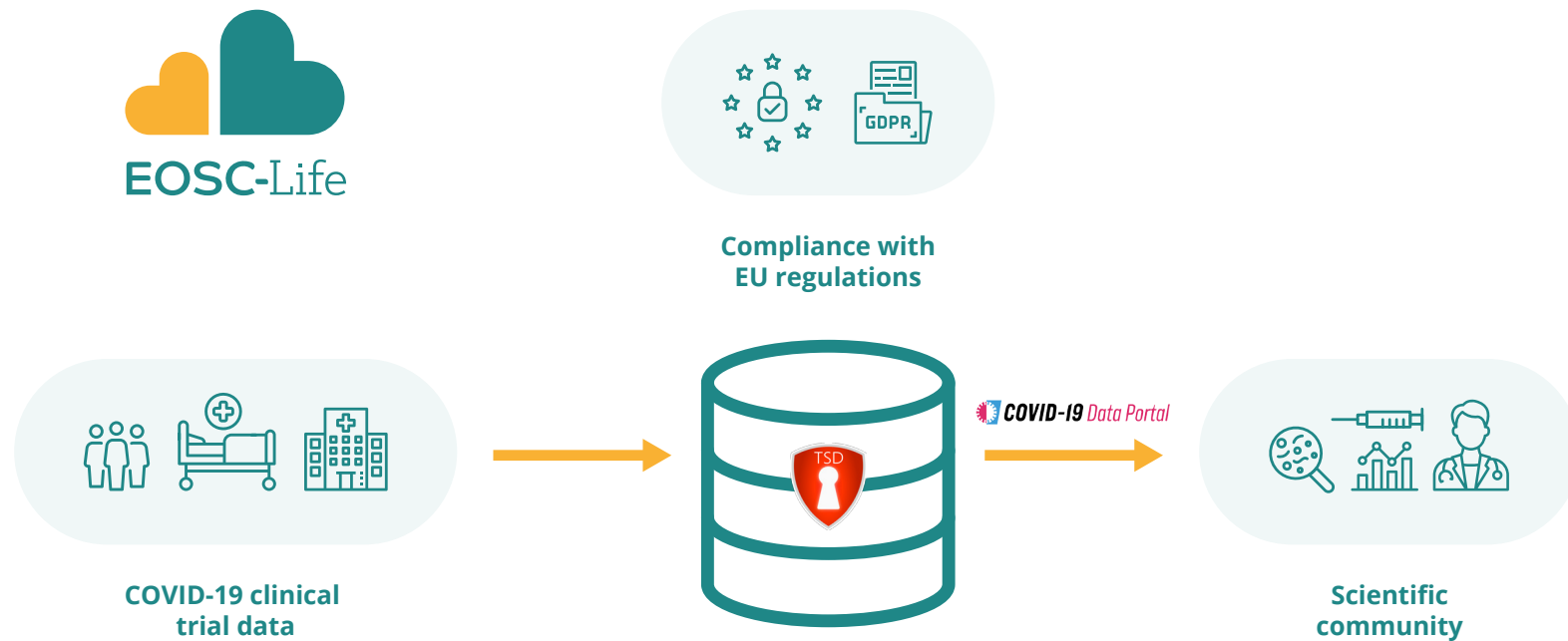
EOSC-Life prepares European life science for a new way of working

By populating EOSC-Life, we are not merely fulfilling a project's goals—we are building tools and training people to make this the new normal for life science data. European scientists should be able to collaborate and reuse data regardless of where they are based. Training programmes, workshops and hackathons help to prepare our users for a new way of working. Open calls and demonstrator projects guide our developments and create examples for other projects—illustrating the capabilities of EOSC-Life.

The legacies of EOSC-Life—parts of the future EOSC—include:

- People: skills, new capabilities and networks
- Experience and processes for consolidated RI service delivery
- Life science data resources, workflows, registries and other services
- Training materials, policies, guidelines (FAIR, Sensitive data)

In the following pages we introduce some of our achievements thus far, and how potential users can take advantage of these to enrich their life science research.

How to benefit from EOSC-Life

EOSC-Life helps research infrastructures evolve the repository infrastructure; for example a particular dataset is being deployed into the cloud and connected to workflows, enhancing the accessibility of mouse model data for researching biology and understanding diseases. EOSC-Life is developing data expertise within the Life Science Research Infrastructures, helping to build a community of experts across the European research infrastructures.

Science Demonstrators

EOSC-Life has funded 8 "Demonstrator" projects, scientific and technical pilot projects that provide concrete scientific use-cases and guide and structure the work done in EOSC-Life. They provide concrete examples of how to build an open digital and collaborative space for biological and medical research. The demonstrator projects represent a broad scope of life science domains and aim to make data, tools and workflows available in the cloud for re-use by the scientific community. The year-long projects have completed and publications and resources are becoming available, including training material and documentation to facilitate uptake in the community.

Clinical Research Metadata Repository

EOSC-Life partners have been working on the development of the Clinical Research Metadata Repository (MDR), including COVID-19 data, allowing researchers to access clinical studies and related data objects. These include, for example, protocols, information sheets and consent forms, data management plans, statistical analysis plans, case report forms, results, publications, descriptive metadata, etc. MDR contributes to making clinical research data from all disease areas FAIR by increasing data Findability. Browse and search metadata on clinical trials—as well as all related documents.

Cloud Data Deployment Call

Launched in January 2021, 8 funded projects will make their dataset / service / data resource newly available in a cloud instance, implementing FAIR metrics.

COVID-19 Clinical Trial Data Repository

During a pandemic there is a need for timely and accurate collection, reporting and sharing of Individual Participant Data (IPD) from clinical trials. IPD should be stored in trustworthy data repositories that have been certified, are subject to rigorous governance, committed to longer-term preservation of their data and compliant with the relevant regulations (e.g. GDPR). This will facilitate re-analyses, secondary analyses and patient-level data meta-analyses. With the help of EOSC-Life, our partners are currently developing such a repository using the TSD platform for sensitive data. The repository will be part of the European *COVID-19 Data Portal*.

Extension of the COVID-19 Data Portal

The COVID-19 Data Portal brings together relevant datasets submitted to major centres for biomedical data, with the aim to facilitate data sharing and analysis, and to accelerate coronavirus research. Within EOSC-Life the COVID-19 Data Portal is extended to mobilise open biomolecular data (500,000 records from the biomolecular and literature domains are available openly to users), to mobilise new SARS-CoV-2 data (currently >160,000 viral isolates with raw sequence data; 75% of the world's data flows through the Data Hubs) and to connect to clinical and epidemiological data.

FAIRsharing.org (standards, databases, policies)

FAIRsharing is a place to discover standards, repositories and data policies and how they interlink. It also features a prototype educational page on standards. The resource serves all disciplines and has been *adopted* by funders, publishers, RDA and other organisations.
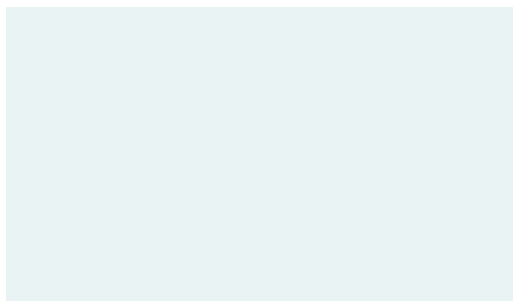
FAIRsharing EOSC-Life Collection

A collection of 100 diverse data resources (each containing thousands of datasets) following FAIR principles, produced by EOSC-Life partners.

Provenance standard for life science data

A provenance model is being developed and its instance is being standardised under ISO 23494 to describe the history of data in life sciences in distributed environments, in order to assess reusability of data for further research and to improve reproducibility of research results. The model aims at documenting the full chain of history of data all the way to its source (biological entity), supporting documenting history of sensitive data (such as genetic or clinical data) and is designed to implement interlinking across institutional boundaries. The standard also supports compliance with the *Nagoya Protocol*.

Packaging tools and workflows for use in the cloud

EOSC-Life is developing expertise in the cloud deployment of software and workflows across all domains of the life sciences. This will help Life Science Research Infrastructures to develop their computational infrastructure and make it FAIR.

WorkflowHub

A registry of scientific workflows to make them more findable, mostly from EOSC-Life partners, many from COVID-related initiatives. The federated registry would support a common API to simplify access for tool developers. An open community of developers and users has been formed and is growing, as the *WorkflowHub Club*. WorkflowHub has already been mentioned in a *Nature Methods* article. Visit WorkflowHub to find and retrieve scientific workflows, or to register your own workflow.

COVID workflows in the WorkflowHub

Due to the COVID-19 pandemic, the focus shifted towards providing tools and workflows to tackle analyses of COVID-related data in an open and reproducible way. The WorkflowHub was accelerated by 6 months and released early to provide a registry for COVID-19 workflows. In collaboration with the global Galaxy community, workflows have been made available on *public Galaxy instances* worldwide.

Tools Collaboratory Roadmap

Our roadmap promotes interoperability across LS RIs and domains and can be used by less mature infrastructures to shape their own computational infrastructures in a more coherent manner.

Metadata specifications for tools and workflows

Community definitions of the metadata required for describing, documenting and registering a Computational Workflow and a Computational Tool were developed and harmonised. The schema has been presented using Schema.org, a web metadata markup standard, as part of the Bioschemas community, which produces subsets of *Schema.org* suitable for the biosciences resources. *Bioschemas* markup enables discovery by search engines and other aggregators such as Google and OpenAIRE.

RO-Crate

RO-Crate is a community effort to formalise packaging of research data with structured metadata, based on *Schema.org*. EOSC-Life has been instrumental in developing the RO-Crate community and specifications, in particular with its specialisation Workflow RO-Crate used by *WorkflowHub* for integrating with workflow systems such as *Galaxy* and continual workflow monitoring with the Workflow Execution service WfExS. RO-Crate is seeing wider interest and adoption across EOSC initiatives, and we are also aligning RO-Crate with FAIR Digital Objects through Research Data Alliance and the DISSSCo SYNTHESYS+ project.

Cloud deployment and cloud resources

EOSC-Life provides access and support to cloud resources for the life science community and offers deployable platforms for workflows. In addition, we establish capabilities and identify suitable cloud providers for secure hosting of sensitive data.

Helpdesk system for cloud deployment and cloud resources

EOSC-Life's cloud deployment and cloud resources work package has set up a helpdesk system that can be contacted by sending an email to *eosc-life-helpdesk@elixir-europe.org*—currently limited to project participants—to provide support for technologies within the project (e.g. *Galaxy*). The helpdesk system follows a *support procedure* to offer consistent answers to all users.

Providing access

We are creating an access and user management system to enable multi-RI applications and workflows that build on existing approaches and support access to sensitive data with their specific requirements.

LS AAI

The Life Science Login enables researchers to use their home organisation credentials or community or commercial identities (e.g. ORCID, LinkedIn) to sign in and access data and services they need across multiple platforms. It also allows service providers (both in academia and industry) to control and manage access rights of their users and create different access levels for research groups or international projects. Piloting has begun.

Connecting users

EOSC-Life is a user-focussed project in which open calls for user research allow the wider life science community to receive support to make data FAIR, share it in the cloud and to support the implementation of large-scale data analysis workflows in the cloud. We are continually aligning our strategy with those of other entities in the EOSC, other cluster projects, and the wider life science community. The collaborative platform and our inclusive community and capacity building will help to foster data science skills in life-science research, within and beyond the research infrastructures.

Digital Life Sciences Open Calls

The 1st EOSC-Life Open Call for projects sharing data, tools and workflows in the cloud generated a huge amount of interest across the scientific community. Project proposals from the first Open Call are currently under review with an expected funding rate of about 10%.

Reach out to our experts during the mandatory maturation phase (deadline 26 May 2021) and apply by 30 June 2021 for the 2nd Open Call's sensitive data track or industry track.

International Collaborations

EOSC-Life is collaborating with the other EOSC Science Cluster projects to contribute to common activities and position papers. We also align with EOSC Governance activities and provide strategic input, such as to the Consultation on the Strategic Research Agenda of the *EOSC Association*. A dedicated "EOSC Stakeholder Interaction Group" (ESIG) discusses EOSC Governance-related topics with representatives from all Life Science Research Infrastructures participating in EOSC-Life.

Informing & Training Users

Training the EOSC-Life community—from staff to end users—on FAIR principles, using sensitive data in research, and the use of our data resources, tools and workflows, is an integral part of our mission

FAIRassist
The FAIRassist tool, part of the FAIRsharing resource, is under development to offer personalised guidance to discover resources, such as data and metadata standards and databases, which should be used to make data FAIR.

Terms4FAIRskills
Terms4FAIRskills is a terminology for the skills, competencies and knowledge necessary to make data FAIR and to keep it FAIR. The terminology can be used to assist with the creation and assessment of stewardship curricula; to facilitate the annotation, discovery and evaluation of FAIR-enabling materials (e.g. training) and resources; and to enable the formalisation of job descriptions and CVs with recognised, structured competencies.

Toolbox for sharing of sensitive data
Sharing sensitive data is a specific challenge within EOSC-Life. For that reason, a toolbox is currently under development, providing pooled information on recommendations, best practices, software tools etc. to researchers who wish to share and/or use sensitive data in a cloud environment in general, and the European Open Science Cloud in particular. The sensitivity of the data may arise from its personal nature but can also be related to intellectual property considerations, biohazard concerns, or the Nagoya Protocol.

Landscape mapping on sharing and re-use of health data
EOSC-Life is currently mapping the national landscapes on sharing and re-using health data to understand the implications for Life Science RIs and their services, especially with regards to data protection, appropriate safeguards (e.g. de-identification techniques), risk-based approaches, data ownership, and conditions for sharing and re-use of health data. The results will allow us to assess EU countries' preparedness for EOSC.

Training Open Calls
Our Training Open Calls offer funding to support training activities and provides expertise and guidance to help awardees organise successful and impactful training. Examples of funded projects are training and mentoring for *Open Life Science* Ambassadors, COVID-19 modelling, and cross-cluster training on modelling workflows with FATES for improving climate models.

Training resources
EOSC-Life has organised several training events for the EOSC-Life community to prepare European life science for a new way of working, and is supporting RIs in transferring their face-to-face training to a remote setting by organising a remote learning series. View upcoming training and past training materials and recordings, such as the online tutorial on *ResOps, cloud native tools and technology for researchers.*

For more information about the project and to stay updated on future achievements, visit **www.eosc-life.eu** and sign up for our bi-monthly **newsletter**.

Follow us on:
🐦 **Twitter**
💼 **LinkedIn**

EOSC-Life