



The European Research Infrastructure
Consortium for Structural Biology

FAIR data: Instruct's approach and current work

Susan Daenke

Instruct-ERIC Hub Coordinator

9th November 2020



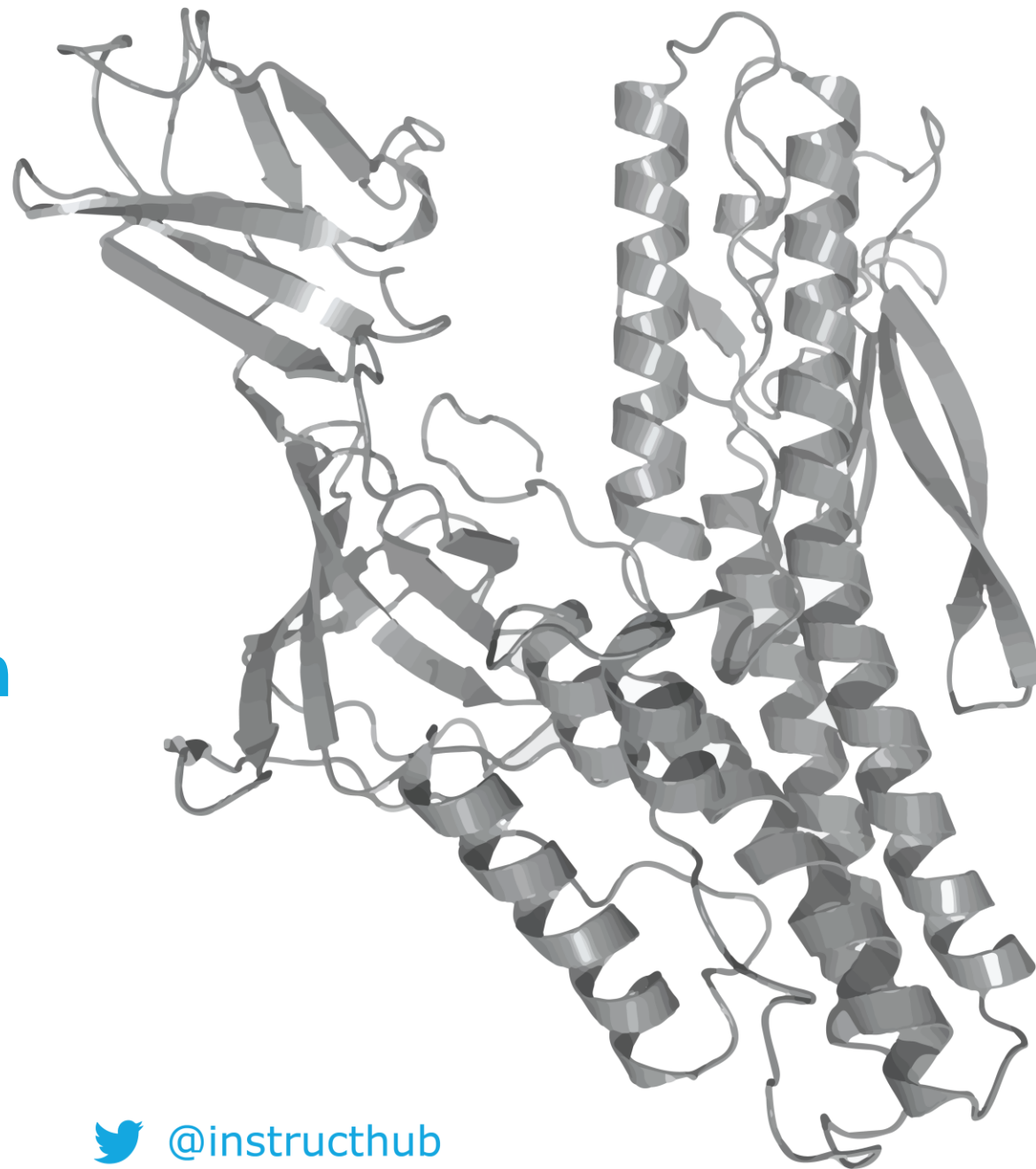
instruct-eric.eu



admin@instruct-eric.eu

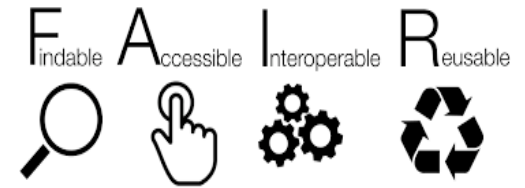


[@instructhub](https://twitter.com/instructhub)



Instruct-ERIC - structural biology facilities for researchers

- Early stages of central coordination towards FAIR

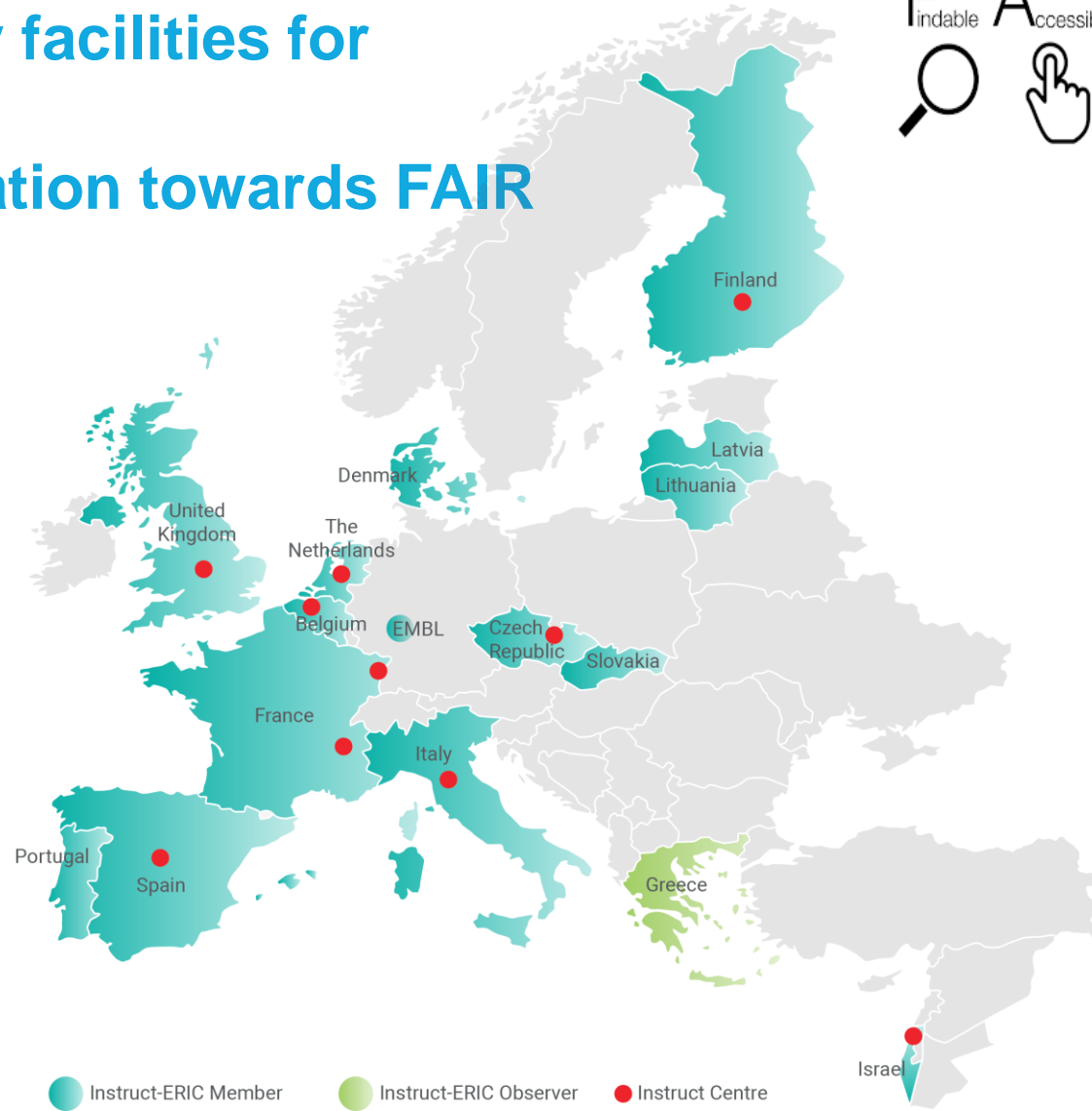


Instruct-ERIC is a pan-European distributed research infrastructure making high-end technologies and methods in structural biology available to users.

Our aim is to promote innovation in biomedical science and operates on a non-economic basis within the scope of the ERIC Regulation.

Instruct-ERIC encourages the use of combined techniques by facilitating and funding visits to experimental facilities, and providing training to researchers learning new techniques.

Instruct operates through 10 Centres, coordinated from the Hub offices in Oxford, UK.



instruct-eric.eu



admin@instruct-eric.eu



[@instructhub](https://twitter.com/instructhub)



First step in FAIR data management

West-Life

(2016-2019)

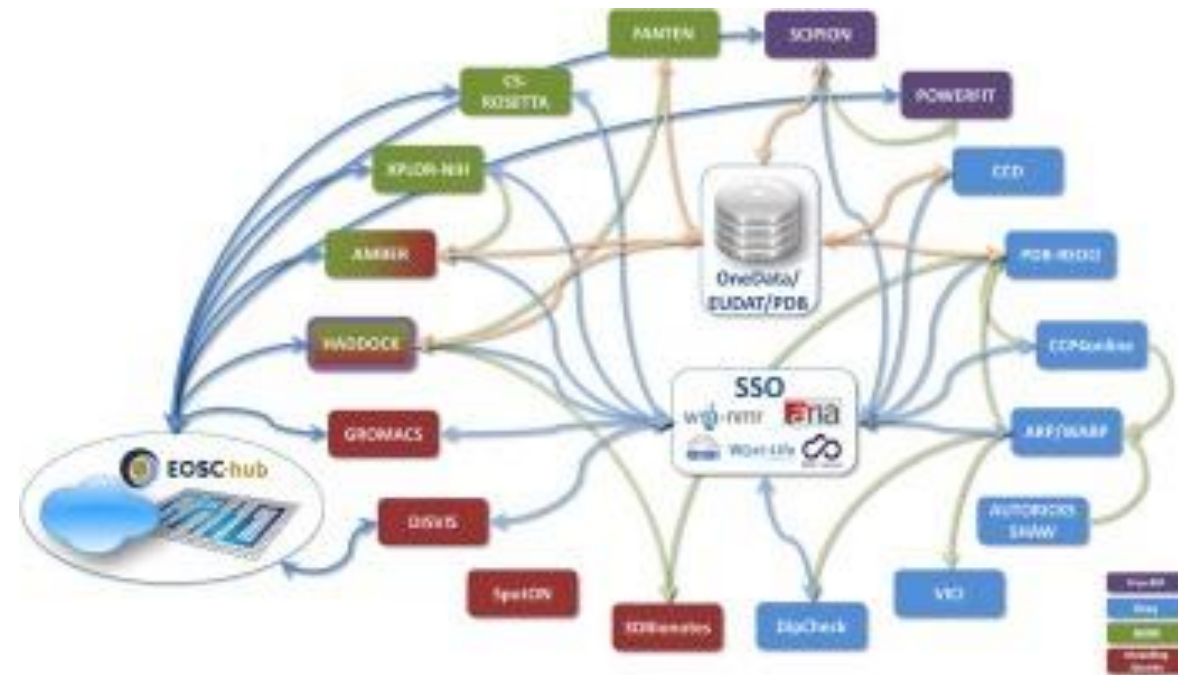
EU funded project to provide data processing and data management services for the international community of structural biologists

supported integrative experimental approaches within the field of structural biology

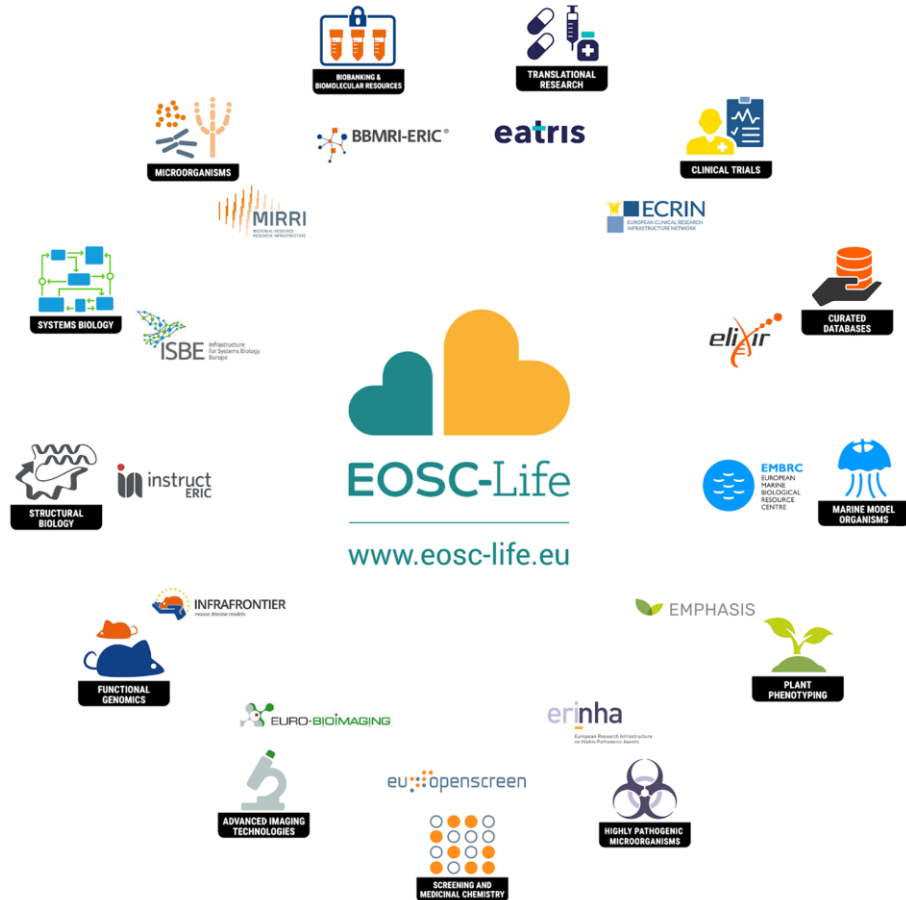
created new pipelines to link these services into more complex higher-level workflows, and added new data management facilities

Included multiple components handling data processing, data management, compute resources, infrastructure for authentication and authorisation, quality assurance and user help.

Virtualised portal to broad range of data management tools



Next step – EOSC-Life: an open collaborative space for digital biology in Europe



- 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

PaNOSC also has interactions with EOSC-Life through CERIC-ERIC, ESS, X-FEL, ILL, EGI, GEANT
Participants in EOSC-Future



instruct-eric.eu



admin@instruct-eric.eu

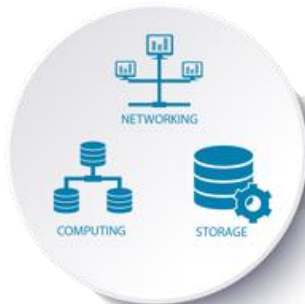


[@instructhub](https://twitter.com/instructhub)



EOSC-Life will shape EOSC for life sciences data needs

European **O**pen **S**cience **C**loud =



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

E-Infrastructure consolidation

+



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

Open Science

+



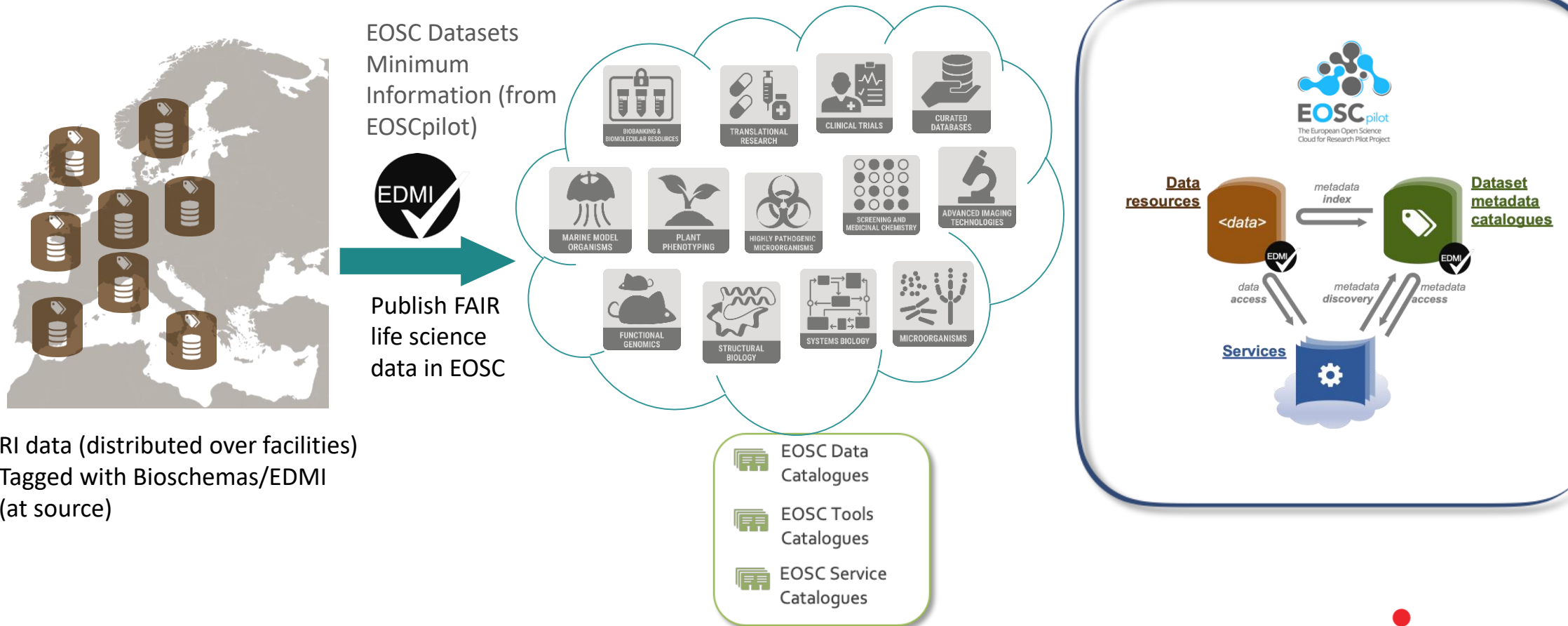
Populate EOSC with the scientific data resources and computational tools from research infrastructures - drive usage by to Europe's 1.7 M researchers

Scientific Communities' content and users

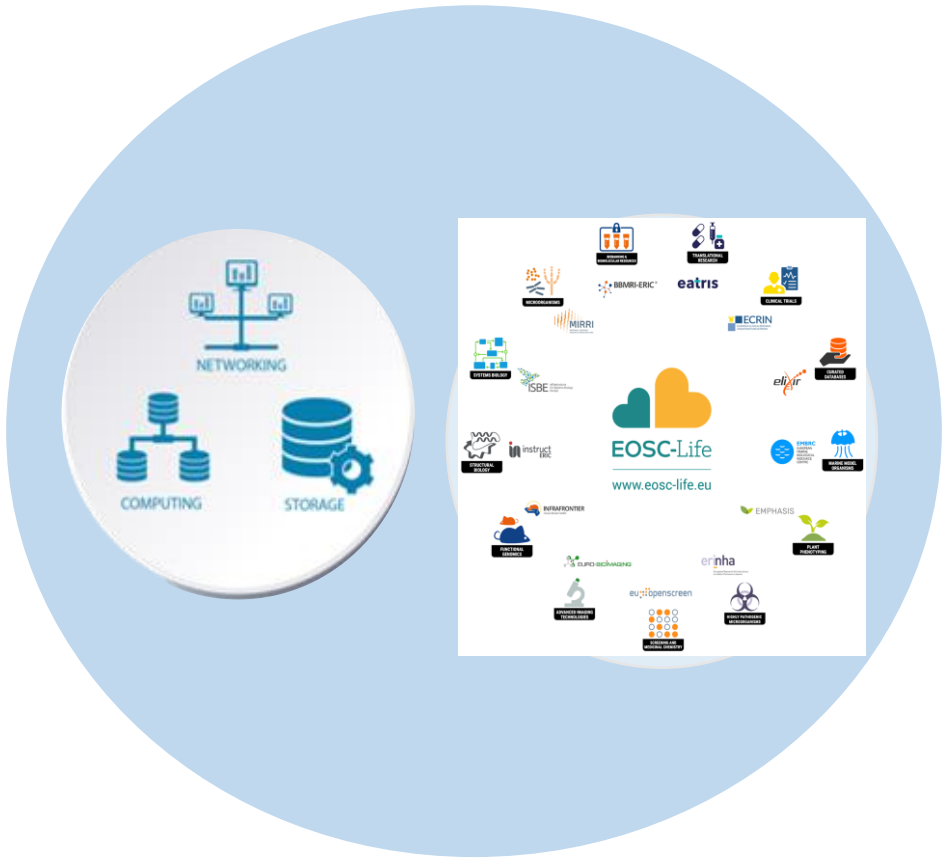


Objective: to publish data and tools for cloud use : create EOSC for the life sciences

- EOSCpilot: EDMI standard for harvesting distributed, FAIR, data (and metadata) and tools
- EOSC-Life: Populate the life-science registries that will feed the EOSC catalogues



Instruct-ERIC co-leads AAI and single sign-on



Co-create and integrate *EOSC federating core*

EUROPEAN OPEN SCIENCE CLOUD

Find service... All services ▾ 🔍 My EOSC Marketplace

Services

All Services 6

CATEGORIES

- Compute 6
- Data management 0
- Networking 0
- Processing & Analysis 3
- Security & Operations 0
- Sharing & Discovery 0
- Storage 4
- Training & Support 1

FILTERS

Research Area ^

Find or choose from the list below

- Interdisciplinary 4
- Engineering and Technology 3
 - Bioengineering and biomedical engineering 1
 - Chemical engineering 1
 - Electrical, electronic and information engineering 1
 - Aerospace engineering 0
 - Civil engineering 0
 - Environmental biotechnology 0
 - Environmental engineering 0
 - Industrial biotechnology 0
 - Materials engineering 0
 - Mechanical engineering 0
 - Nano-technology 0
 - Other engineering and technologies 0
- Humanities 2
 - Arts 1
 - Philosophy, ethics and religion 1

Services 6

Active filters
Related Infrastructures and platforms: EOSC-Life ✕ Clear all filters ✕

Found 6 results Sort by: by name A-Z ▾

CSC ePouta

Secure and cost-effective cloud computing for processing sensitive data

Provided by: CSC
Research area: Interdisciplinary, Arts, Electrical, electronic and information engineering, Social Sciences, Philosophy, ethics and religion
Dedicated for: Researchers, Research organisations, Research group, Providers

Add to comparison

Embassy Cloud

EMBL-EBI's OpenStack cloud infrastructure co-located with their global life-science data resources and bioinformatics services and tools.

Provided by: EMBL-EBI, European Bioinformatics Institute
Research area: Biological sciences
Dedicated for: Business, Providers, Research group, Research organisations, Researchers

Add to comparison

MetaCentrum Cloud

Czech national scientific cloud

Provided by: CESNET
Research area: Engineering and Technology, Humanities, Interdisciplinary
Dedicated for: Researchers, Research organisations, Research group

Add to comparison

Rahti Container Cloud



instruct-eric.eu



admin@instruct-eric.eu

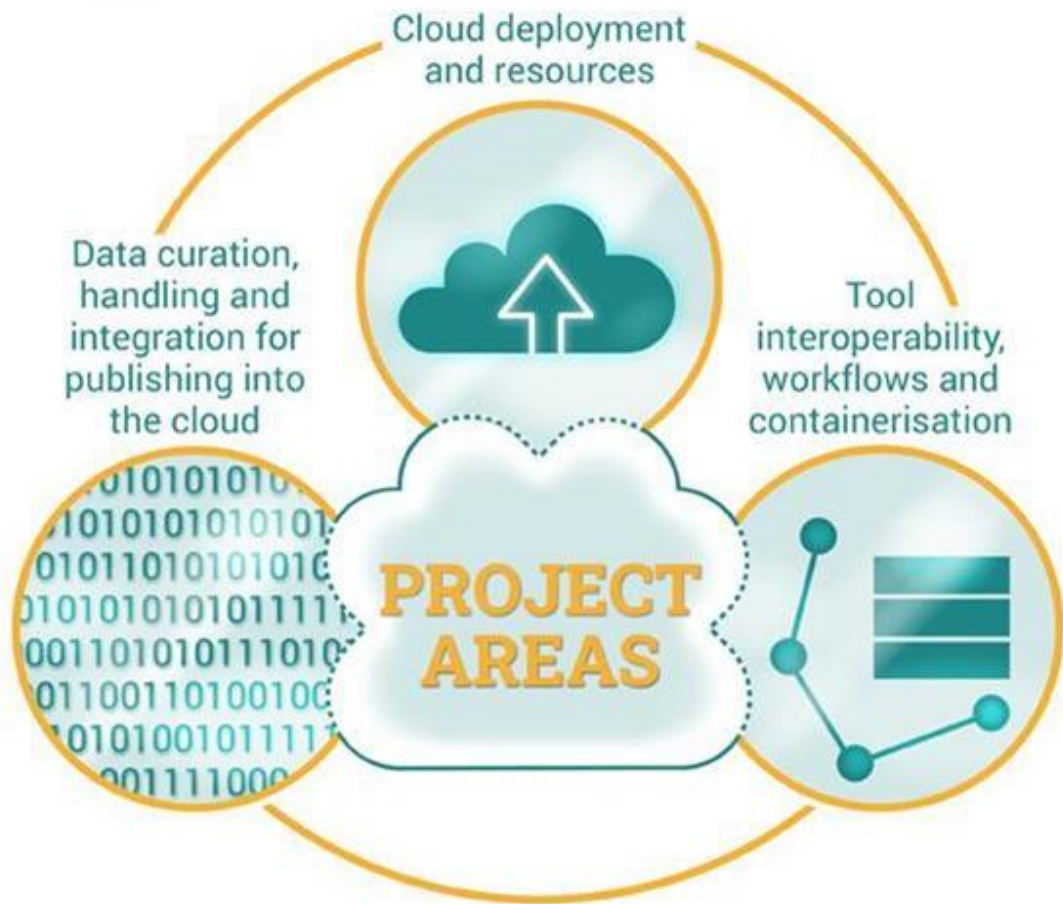


@instructhub



DIGITAL LIFE SCIENCES OPEN CALL

A European Open Science Cloud (EOSC-Life) call for projects sharing data, tools and workflows in the cloud



THE DIGITAL LIFE SCIENCES OPEN CALL OFFERS:

Funding for project team: roughly 1 full-time salary for project duration (1 year)

Training and technical expertise

Help cloudifying your data, tools and/or workflows allowing access to the greater scientific community

APPLY TO THE OPEN CALL

- Visit www.eosc-life.eu/opencall to review details of the call
- Contact our experts to discuss feasibility
- Submit application via ARIA

Instruct has 9 use cases to test processes to make data resources adhere to FAIR principles



WHY EOSC?

A vast amount of data is produced, processed and analyzed daily in the life sciences. The European Open Science Cloud (EOSC) aims to **make data, tools and analysis workflows more findable, accessible, interoperable, and reusable (FAIR)** by the life science community through the **cloud deployment** of these resources.

**APPLICATION DEADLINE:
22 DECEMBER
2020**



instruct-eric.eu



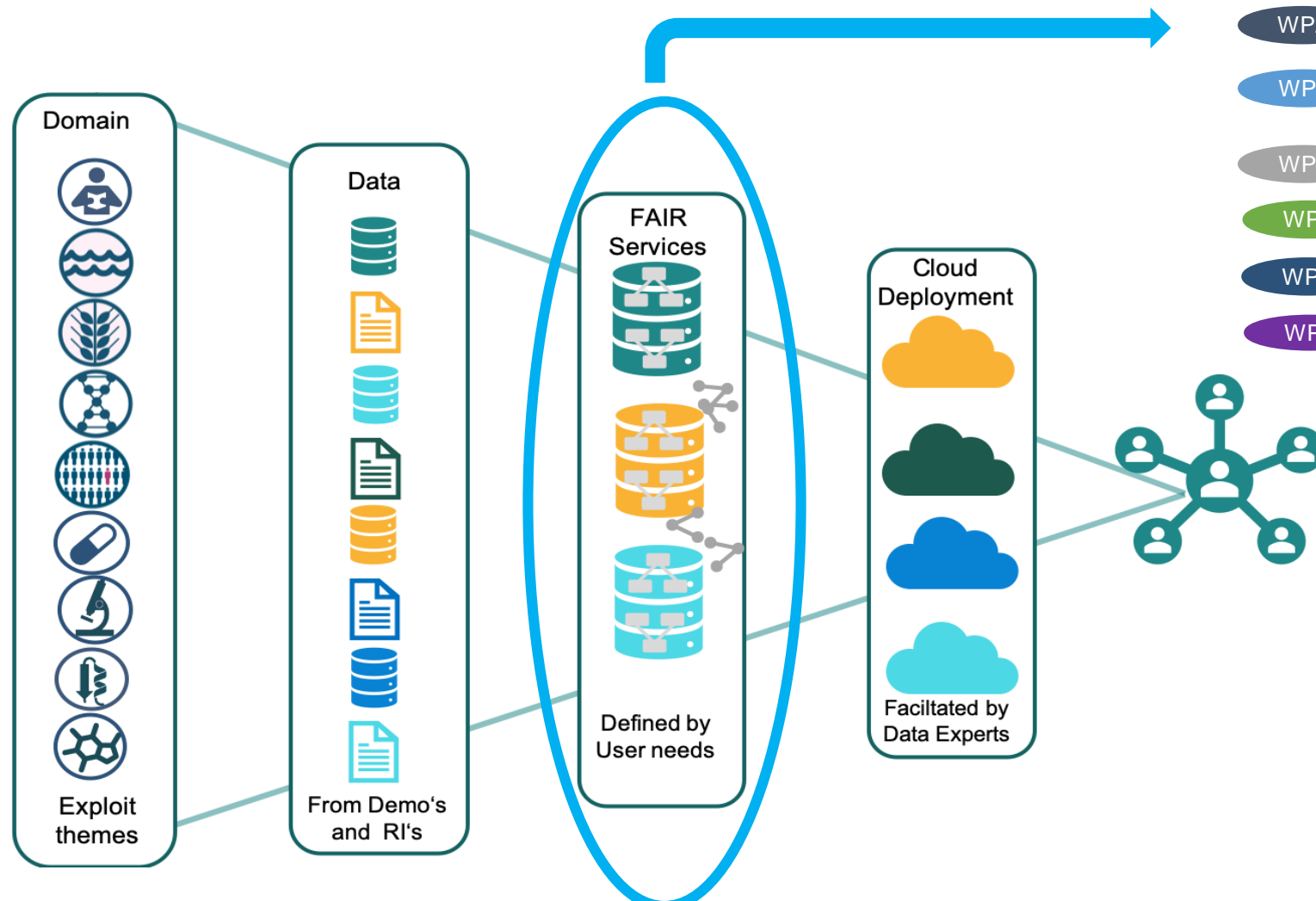
admin@instruct-eric.eu



[@instructhub](https://twitter.com/instructhub)



Publishing FAIR RI data resources in EOSC



Label	Definition
consumed_by	data feeds into analyses
advised_by	data experts advise demonstrators
defines_need	defines ethics needs enabling implementation
implements	outputs implemented
deploys_to	deploys to cloud
documents	Documents for training purposes

WP2: Tools collaboratory; integration and analysis workflows
WP3: Managing and supporting the call applicants
WP4: policies, specifications for biological and medical data
WP6: FAIRification and provenance services for data outputs
WP7: provide set of integrated cloud resources to support FAIR RI data resources, workflows



Application to structural data - why do we need FAIR and interoperable data

Structural data: some data lost at source:

- diffuse scattering for diffraction data
- Imperfect data processing → impaired resolution

Increasing data volumes pose problems for local repositories: (how long to keep primary data; how to archive)

- SBGrid is a possible solution – all stored data assigned doi
- Other smaller repositories emerging
- EOSC

Imperfect structure solutions need review:

- Difficult to correct → impact if used for future research
- IUCR now asks authors to provide permanent link to raw datasets and processed data
- Allows reuse of original data to rectify errors and add higher quality structures to PDB (PDB-REDO)

Primary structural data is well annotated

Metadata is more problematic for FAIR compliance (no agreed ontology, often incomplete; use of PDBx/mmCIF files help by allowing incorporation of more metadata)



What is available to help in this process?

FAIRassist (fairassist.org): run from University of Oxford – provides a list of resources for the assessment and evaluation of data against FAIR principles;

FAIRsharing.org: catalog of metadata standards, inter-related databases and data policies (COVID-19 response required agility and extension of metadata cataloguing)

FAIRplus (fairplus-project.eu): developing tools and guidelines for making LS data FAIR

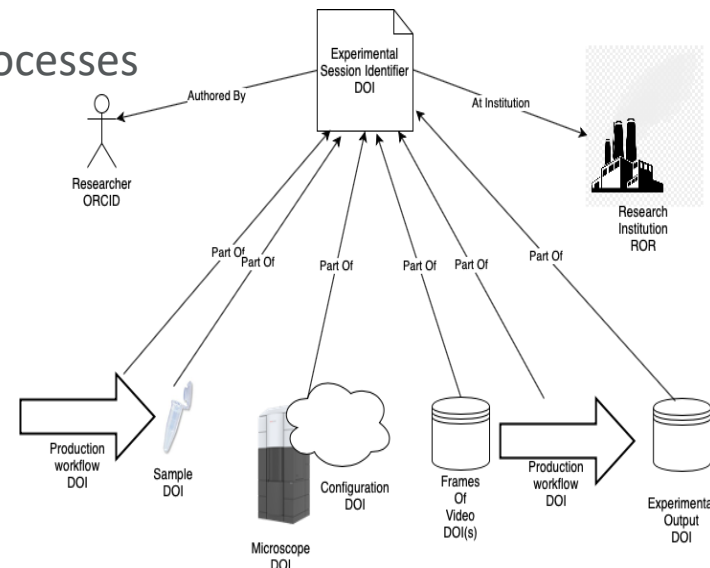
What is else is Instruct doing?

EOSCPilot: cryo-EM workflows enabling user to obtain raw and preprocessed data with a file linking to the data and analysis workflows which enable reproducibility of processing and deposition in defined DB

Instruct is surveying each of its Centres to identify processes already in place to make data FAIR

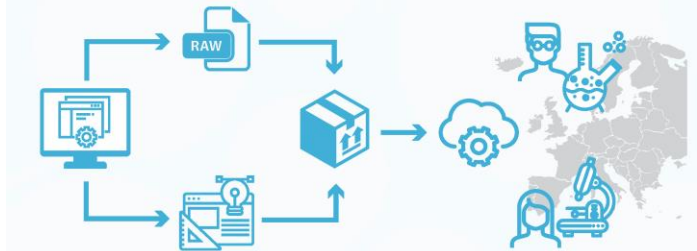
FREYA: aims to extend the infrastructure for persistent identifiers (PIDs) as a core component of open research, in the EU and globally. Build a PID graph to aggregate all citations for a research object (publications, data, software, samples, reagents)

Instruct is planning a model to produce a 'research bundle' that assigns a doi to all experimental objects which remain associated with all other components of the bundle



Brief overview

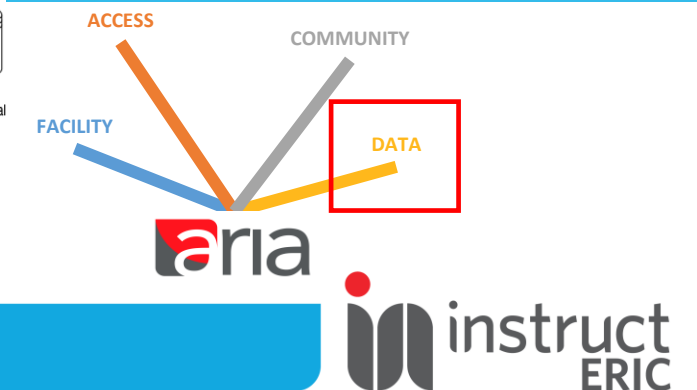
CryoEM aims to improve reproducibility of their work using image processing workflows through the production of a Scipion workflow file that describes their image processing steps. This allows full reproduction of the same results when this data is deposited in public databases. In this way, cryoEM research becomes more transparent and traceable pursuing the spirit of public Open Science.



Objectives

- Enable users of a representative subset of major CryoEM Facilities in Europe to bring back raw and preprocessed data, and a file linking to the acquired data and the analysis workflows.
- The file will contain detailed information enabling the reproducibility of processing steps, be ready and accepted to be deposited in CryoEM major databases, and be easy to browse and analyze over the Web.

EOSCPilot.eu has received funding from the European Commission's Horizon 2020 research and innovation programme under the Grant Agreement no.739563.



For more information about Instruct-ERIC, visit the Instruct-ERIC website.

instruct-eric.eu

To stay informed of the latest opportunities and open calls from Instruct, register for an ARIA account at:

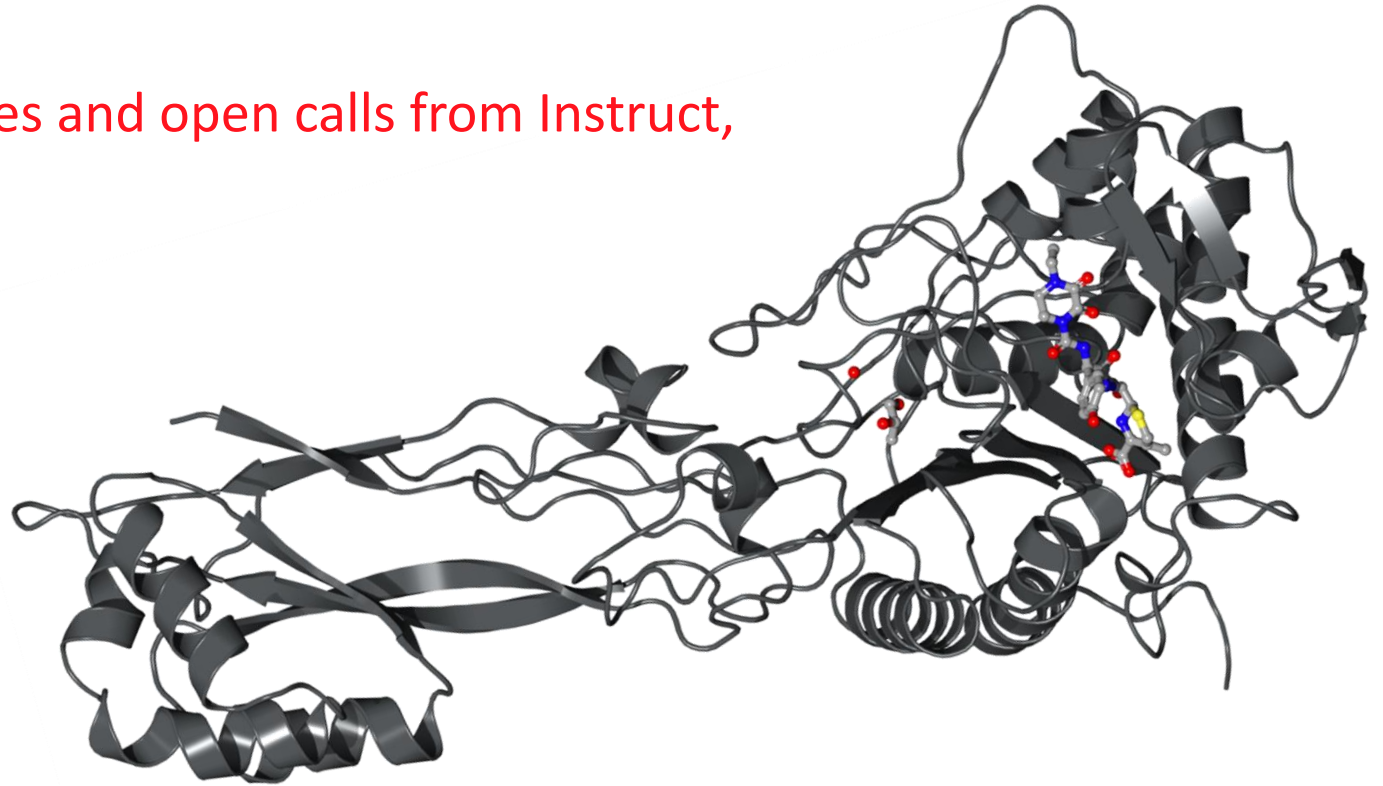
instruct-eric.eu/register

Acknowledgements

Instruct-ERIC is funded by contributions from its members.

West-Life and EOSC-Life are funded by the European Union under the H2020 programme:

Thanks go to these projects for content and images for this presentation



Thank you for your attention



instruct-eric.eu



admin@instruct-eric.eu



@instructhub