

EOSC-Life

Open Science Cloud and
Infrastructure for data in
clinical research

Creating EOSC for
the life sciences

Ludek Matyska

Institute of Computer Science
Masaryk University
Brno
Czech Republic



EOSC-*Life*

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

- Publish FAIR life science data resources in EOSC
- Create an ecosystem of innovative life-science tools in EOSC
- Enable ground-breaking data driven research in Europe by connecting life scientists to interoperable European clouds via open calls for participation



EOSC-Life Consortium

- 13 ESFRI Health and Food Research Infrastructures
- 46 Partners and 17 linked 3rd parties
- Sourcing e-Infrastructure services from EOSC (Cloud, AAI,...)
- 4 year Project, 24M€
- Started 1st March 2019



• biobanks



• curated databases



• marine model organisms



• systems biology



• Plant phenotyping centres



• translational research



• mouse disease models



• screening & medicinal chemistry



• microorganisms



• clinical trials



• structural biology



• biological/medical imaging



• Highly Pathogenic Agents



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087

EOSC Life Workpackages

WP	Title
WP1	Publishing FAIR RI data resources in EOSC
WP2	Tools Collaboratory: Deployment of life-science data integration and analysis workflows in EOSC
WP3	Demonstrators and Open Calls for User Projects
WP4	Policies, specifications and tools for the management of data for biological and medical research
WP5	User management and access services
WP6	FAIRification and provenance services
WP7	Cloud Deployment Services
WP8	International Impact, Innovation and Sustainability
WP9	People and Training
WP10	Dissemination and Outreach (Stakeholder event)
WP11	Project Management (Kick Off + AGMs)



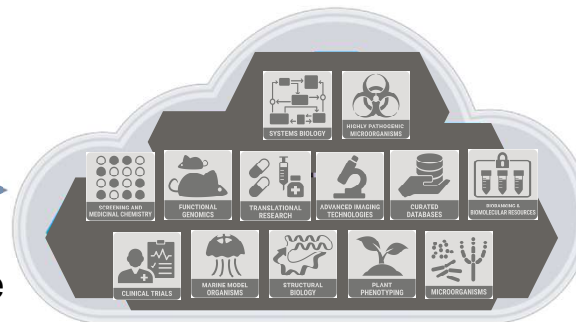
Establish EOSC-Life by publishing data and tools for cloud use

- Health and Food RI are **distributed with** many hundred partner facilities across Europe
- Health and Food RI have large community of tools developers across Europe

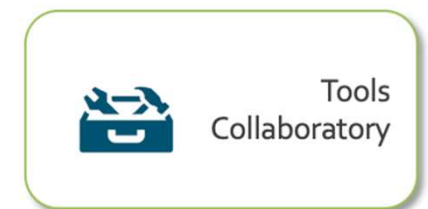


RI data (distributed over facilities)

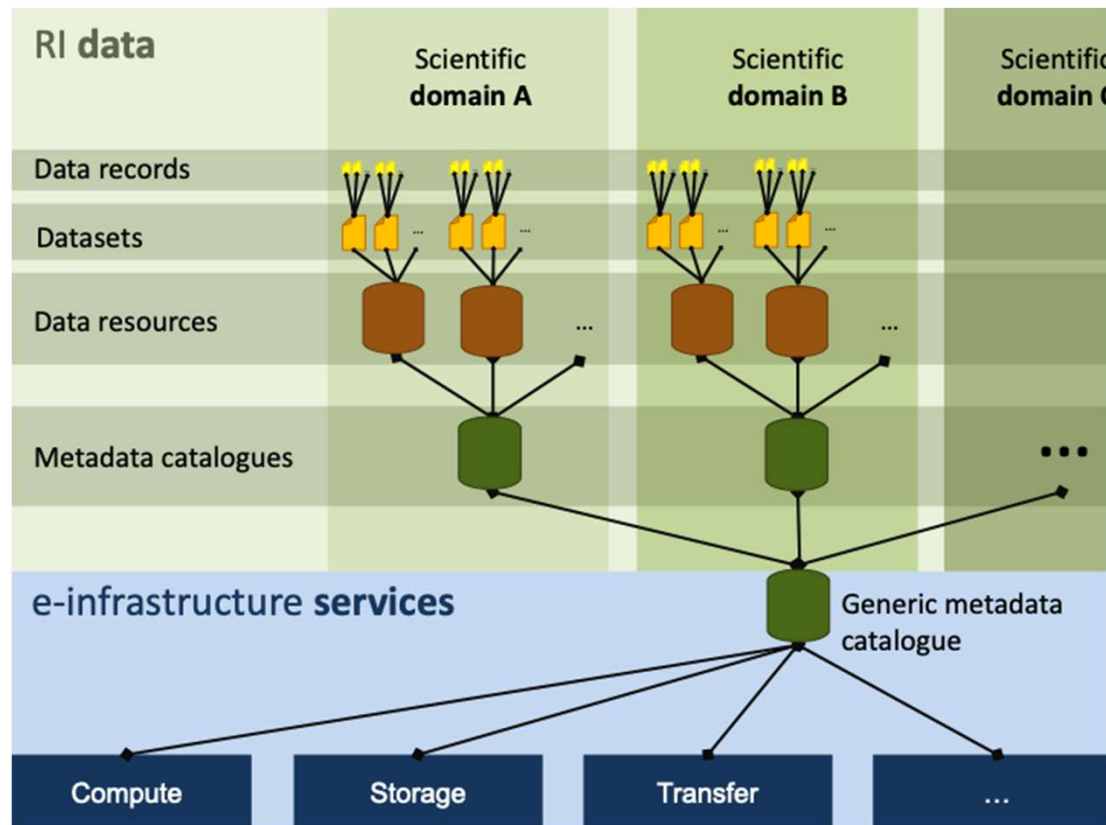
Publish FAIR life science data in EOSC



Ecosystem of innovative tools in EOSC



Creating EOSC for the life sciences



EOSC_{pilot}
The European Open Science
Cloud for Research Pilot Project



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087

EOSC-Life: Cloud Architecture

Cloud Activities

WP1: Data Resources

WP2: Tools

WP3: Demonstrators

WP9: Training

Platforms

Managed Platforms

Container Orchestration
GA4GH Workflows
Appliances

Galaxy
Cloud Portal

Training
Helpdesk

Unmanaged Platforms

Services from WP1
Workflows from WP2
Demonstrators from WP3

EOSC

Pledged Cloud

EMBL-EBI
ELIXIR DE (de.NBI)
ELIXIR CZ (MU)

PAYG Cloud

CSC
Other RIs....

EOSC

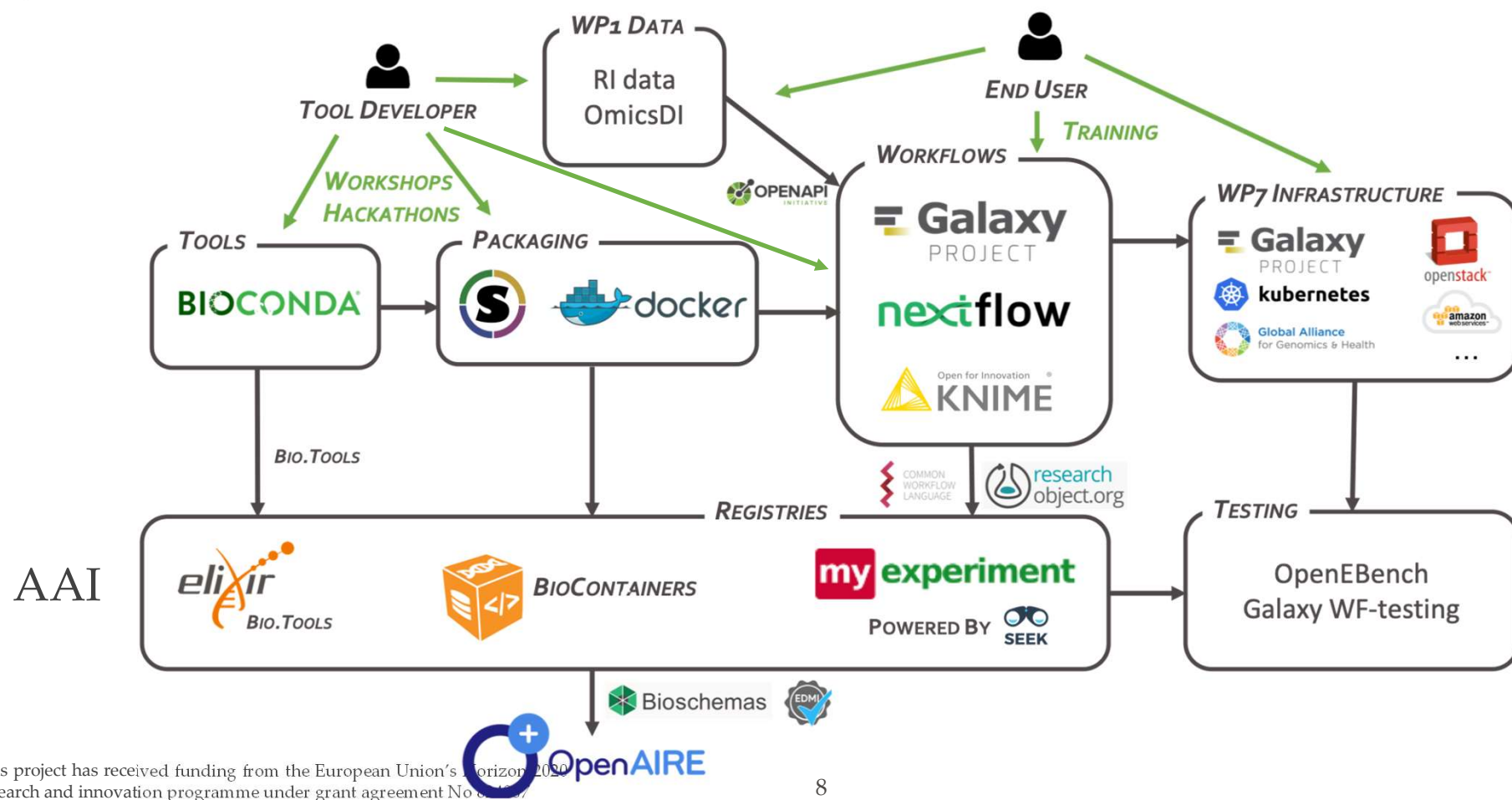
Others...

Commercial Frameworks

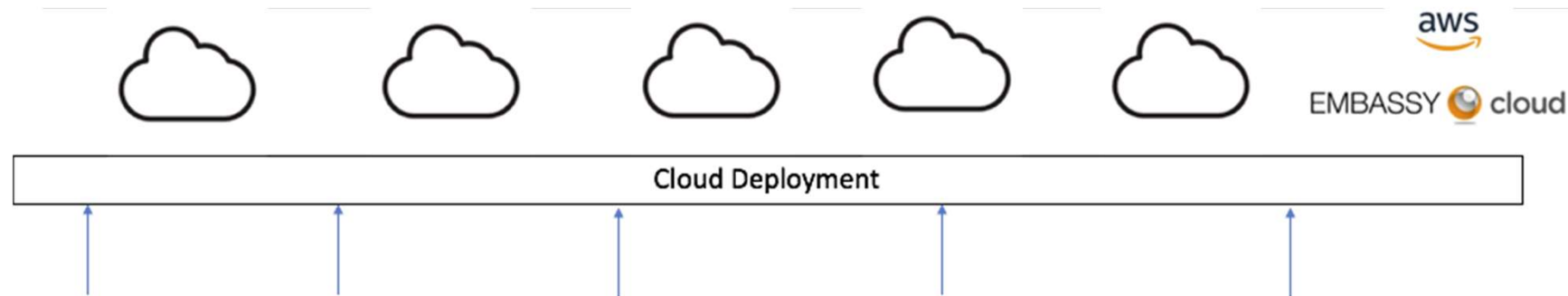
GEANT Cloud Framework
Helix Nebula Science Cloud
OCRE



Tools and Workflows in Containers



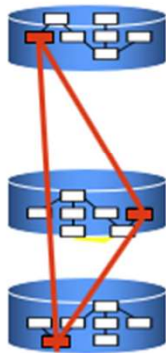
Rich landscape of cloud deployment scenarios



Data



FAIR services



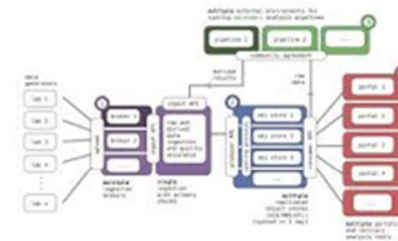
Cloud Workflows



Cloud Mirrors

- UK (Sanger Institute) - YOU ARE HERE!
- US West (Amazon AWS) - Cloud-based mirror on West Coast of US
- US East (Amazon AWS) - Cloud-based mirror on East Coast of US
- Asia (Amazon AWS) - Cloud-based mirror in Singapore

Native Cloud Development/Deployment



European Health Research and Innovation Cloud

Data sharing and reuse

- Protocol design
- Site selection
- Patient selection
- Informed consent
- Data from cohorts / registries
- Electronic health records, hospital data
- Electronic data capture
- Data from national databases
- Data sharing for peer-review, meta-analyses, re-analyses, secondary use

Personalised Medicine

- ❖ High-throughput –omics / imaging data
- ❖ Cohort integration
- ❖ Multimodal data management
- ❖ Stratification algorithms



Clinical / medical research in a conflicting environment

Protection of personal data

- EU General Data Protection Regulation (GDPR) 2016/679

vs.

Open Science : FAIR

- clinical research data sharing : transparency, reproducibility,
- data and biosample reuse in clinical research
 - ❖ reuse of research data/samples for research purposes
 - ❖ reuse of health data/samples for research purposes



GDPR applies to personal data

US HIPAA: 18 Items to remove to de-identify health data: (A) names, (B) geographic subdivisions smaller than a state, (C) all elements of dates (except year) for dates that are directly related to an individual, including birth date, (D) telephone numbers, (L) vehicle identifiers, (E) fax numbers, (M) device identifiers and serial numbers, (F) email addresses, (N) Web Universal Resource Locators (URLs), (G) social security numbers, (O) Internet Protocol (IP) addresses, (H) medical record numbers, (P) biometric identifiers, including finger and voice prints, (I) health plan beneficiary numbers, (Q) full-face photographs and any comparable images, (J) account numbers, (R) any other unique identifying number, characteristic, or code, (K) certificate/license numbers.

EU GDPR : pseudonymized data are personal data, anonymized data are not

No clear boundaries : data minimization, risk of re-identification -> **continuum**

Divergent implementation of GDPR: different legal basis in EU countries



EOSC-Life: Management of data for biological and medical research

- EOSC-Life will address **policies and specifications** for
 - the storage, processing, access, sharing and reuse of biological and medical data for research purposes, **with a special focus on sensitive data**
- Essential step to build the European Health Research and Innovation Cloud
 - to handle health data and health research data in a transnational environment



EOSC-Life: Management of data for biological and medical research

- Promote and improve pan-European policies, standards and risk analyses to ensure full regulatory compliance of data within the EOSC, especially sensitive or classified data.
- Propose and improve standardized solutions for working with multi-source (partitioned) environments, where datasets are not directly available for integration and in many cases not even metadata is disclosed in detail.
- Propose solutions enabling secure federated storage, access and sharing of sensitive research or health data.
- Promote integration / knowledge of privacy-enhancing technologies.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087



Opportunities

- Extensive international and cross European Research Infrastructures collaboration
- Targeting (all) aspects of the data processing
 - European Health research and Innovation Cloud
 - Including work with **the sensitive clinical data**
- Czech Republic has a strong role in the project
 - **Co-leadership in two workpackages** (ELIXIR/WP5 and BBMRI/WP6 representatives)
 - Presence/involvement across the project
- EOSC-Life provides an excellent opportunity to work in a pan European environment towards the safe and efficient environment for the storage and analysis of sensitive clinical data in its clean or pseudonymized form



Thank you

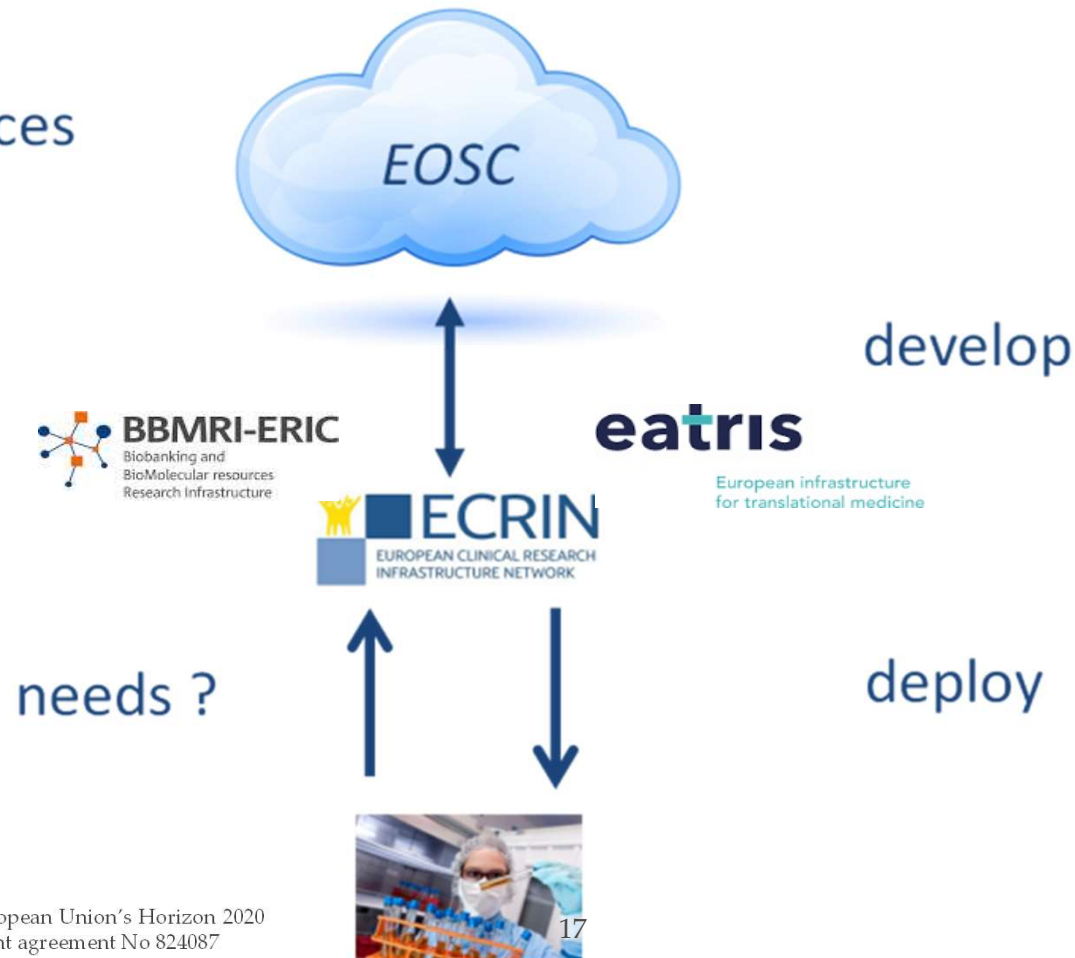
Questions?



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087

EOSC-Life: medical infrastructures in the development and deployment of data services

data services



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087