FAIR-IMPACT
Expanding FAIR Solutions across EOSC

Synergies with EOSC4Cancer

HORIZON-INFRA-2021-EOSC-01-06 101058427

Carole Goble (UNIMAN)
Stian Soiland-Reyes (UNIMAN)

*Thanks to Romina Royo (BSC),
Raquel Villodres Toledo (BSC),
Salvador Capella-Gutierrez (BSC)*

FAIR-IMPACT kick off, Amsterdam, 2022-06-28

EOSC4Cancer
A European-wide foundation to accelerate Data-driven Cancer Research

EOSC4Cancer as provider of the *infrastructure for the exploitation of cancer data* for the **EU Cancer Mission** (10.2777/989951)

EOSC4Cancer brings together comprehensive **cancer centres**, **research infrastructures**, leading **research groups**, and major **computational infrastructure** across Europe to make the exploitation of the data possible.
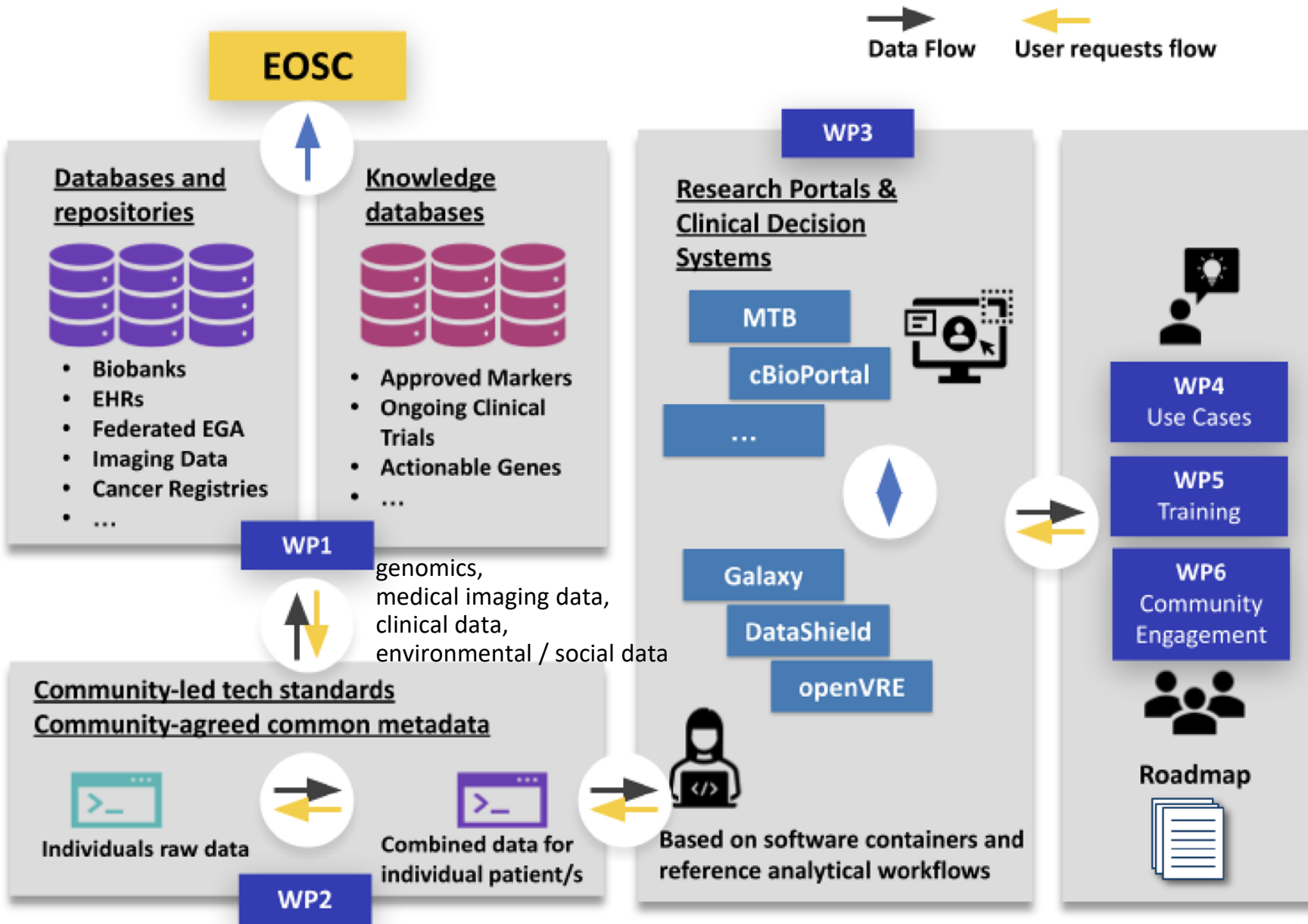
EOSC4Cancer will prepare **EOSC services** for cancer research and enrich EOSC with **data, tools** and **services** from the cancer community.

*) Provisional logo

Projected start: Autumn 2022 (30 months)   Consortium: 34 partners, ~800 PM

Partners and associates of EOSC4Cancer

- **Barcelona Supercomputing Center** (BSC), ES (*coordinator*)
  Contact: romina.royo@bsc.es
- BBMRI-ERIC, AT
- ELIXIR / EMBL-EBI, DE/UK
- DTL Projects, NL
- CSC, FI
- University of Oslo (UiO), NO
- Spanish National Cancer Research Center (CNIO), ES
- The French National Centre for Scientific Research (CNRS), FR
- Palacký University Olomouc, CZ
- Netherlands Cancer Institute (NKI), NL
- The University Medical Center Groningen (UMCG), NL
- The Centre for Research & Technology, Hellas (CERTH), GR
- Italy's National Research Council (CNR), IT
- Fundació Hospital Universitari Vall d'Hebron (VHIO), ES
- Empirica, DE
- Eatris ERIC, NL

- Centre for Genomic Regulation (CRG), ES
- German Cancer Research Center (DKFZ), DE
- INSTRUCT-ERIC, UK
- EU-OpenScreen ERIC, DE
- Institute of Molecular Genetics of the Czech Academy of Sciences (IMG), C
- Albert-Ludwigs Universitaet Freiburg, DE
- European Clinical Research Infrastructure Network (ECRIN), FR
- Euro-Bioimaging ERIC, FI
- Erasmus University Medical Center (ERASMUS MC), NL
- University of Turin (UNITO), IT
- InfraFrontier, DE
- Karolinska Institutet (KI), SE
- Masaryk University, CZ
- European Cancer Patient Coalition (ECPC), BE
- Institute for Research in Biomedicine (IRB) Barcelona, ES
- University of Bordeaux (UBx), FR
- Lygature, NL
- The University of Manchester (UNIMAN), UK

EOSC4Cancer Concept

1. Use-cases trigger questions in the stages of the **Cancer Patient Journey**:
   →Prevention
   →Diagnosis
   →Treatment

2. **Cancer analysis systems** will be engineered to handle particular questions

3. Systems request (via standard interfaces) necessary **data** and **methods**

4. Required data from heterogeneous source will be **harmonised** and made **accessible** under the proper *legal conditions*

FAIR-IMPACT

Enable **storage**, **access, sharing**, **analysis** and **processing** of research data
and other digital **research objects** from basic and clinical cancer research.

> … **registry** of available data … **metadata** descriptions and access mechanisms … **Data discovery** services across data sources. … **Linked data** of European research infrastructure …

Mobilise, **interconnect** and **interoperate datasets** relevant to cancer research.

> … **discovery**, **access** and **reuse** of cancer data (*genomics, medical/imaging data, environmental/social data, clinical data*) … **EOSC-Life** .. **EOSC open science** practices, ... open infrastructure of **data processing**, **workflows**.. distributed heterogeneous environments, **quality management** and **provenance standards**, provide **reproducibility** assurance. … incorporate **FAIR and open data** in cancer research, diagnosis and treatment.

Make cancer research **data** and **analysis systems** accessible to *clinical scientists* in cancer **analysis portals**

> … cancer-related knowledge resources and **catalogues** integrated into **data analysis** and **visualisation platforms** … **without asking them to adopt new systems** … open and extensible **metadata framework** … comprehensive **indexing** of the cancer resources based on **mappings** … **data and analysis framework** for researchers, policy-makers and patients/survivors … **trust in results** and **credit** to **data submitters**, **workflow contributors** and **resources**. … **reproducible** and **trusted data analysis processes … virtual research environments**

Define reference **workflows** and integrate **digital tools**, including **Machine Learning** ones,
for the analysis of cancer data in the cancer analysis portals

> Access data using **open standardised methods**, … Infrastructure for data analysis composed of **Virtual Research Environments** and **workflow** managers (**Galaxy**) … **software containers** developed following **ELIXIR Research Software Best Practices** and the **FAIR principles for research software** … **sustainability** of the EOSC4Cancer software and guidelines

Contribute to **European Health Data Space** (EHDS), **European Open Science Cloud** (EOSC) and the **Cancer Mission**:

> … Guidelines/procedures for **FAIR data management and access.** ... European **COVID-19 Data Platform** and **BY-COVID** ... "*Cancer Data view*" in *ELIXIR Research Data management Kit* (**RDMkit**) that captures good practices, training resources and connects to data experts across Europe.

FAIR-IMPACT

Enable st...s from basic
and clinic...

... reg... ...ked data of
Europ...

Mobilise,...

... dis... ...ife .. EOSC open
scien... ...ement and
prove... ...atment.

Make car... ...alysis portals
... car... ...them to adopt
new s... ...data and
analy... ...w contributors
and r...

Define re... ...of cancer data
in the ca...

Acces... ...d workflow
mana... ...r research
softw...

Contribu... ...Mission:
... Gu... ...Data view" in
ELIXI... ...ss Europe.

Nominated **FAIR contact**:
Salvador Capella-Gutierrez
<salvador.capella@bsc.es>

FAIR-IMPACT and EOSC4Cancer synergies #1

Real use cases of **FAIR data reuse**

… across multiple *European Research Infrastructure Consortia* (ERIC)

→ T5.4 FAIR data and code in FAIR-enabling TDRs

→ T7.2 Engagement at EU/national/institutional level

→ T7.3 Open calls & Cascading grants

→ T7.4 Key impact pathways, adoption & implementation stories

Interoperable data access for **federated** computational systems

→ T6.1 Semantic/technical interoperability across domains

→ T6.3 Interoperability within EOSC

FAIR-IMPACT and EOSC4Cancer synergies #2

Workflow/software **attributions**

→ T4.3 Standard metadata for research software

→ T5.2 FAIR metrics for research software

FAIR **cataloguing** platform w/ harmonization/mapping and *metadata interoperability*

→ T4.2 Semantic artefact lifecycle & catalogues

→ T4.5 FAIR Semantic Artifacts in data repositories

→ T6.2 Legal/organizational interoperability

Training/outreach collaborations (FAIR, **RDMkit**)

→T2.2 FAIR Implementation Framework

→T2.3—T2.5 Engaging & supporting adoption

FAIR-IMPACT and EOSC4Cancer tech

Technological overlaps

FAIR – mostly Accessibility and Interoperability

Metadata mappings

RO-Crate for packaging research outputs with metadata

FAIR Digital Objects (FDO)

FAIR Computational Workflows / FAIR software

Missing in action:

Persistent Identifiers (PID) → T3.2.3/UKDS

BioDT – biodiversity digital twin HORIZON-INFRA-2021-TECH-01-01 101057437 https://biodt.eu/
FAIR workflows, FAIR Digital Objects, semantic mappings

BGE – Biodiversity Genomics Europe HORIZON-CL6-2021-BIODIV-01 101059492
FAIR Data infrastructures, PIDs, RO-Crate

AgroServ HORIZON-INFRA-2021-SERV-01 101058020
FAIR Data management, FAIR knowledge hub, RO-Crate for plant sciences

EuroScienceGateways HORIZON-INFRA-2021-EOSC-01-04 101057388
Reproducible FAIR Digital Objects, FAIR workflows

BioFAIR (UKRI Infrastructure Fund) UK National investment to build a
FAIR Data and Methods BioCommons: *FAIR workflows, Reproducible FAIR Digital Objects - RO-Crate, FAIR Data management, PIDs*