# eosc | cancer

# D7.2 Data Management Plan (DMP)

Main author(s):

| | |
|---|---|
| Ilaria COLUSSI | BBMRI-ERIC |
| Kurt MAJCEN | BBMRI-ERIC |

**Abstract**

This document provides information as to what the relevant project components and repositories are (from each WP) and how they are managed. It is developed by WP7 as an integral part of the governance structure, and oversees the quality assurance of the project outcomes and alignment with FAIR data principles. It also takes the necessary restrictions into account (e.g. patient/clinical data, requirements with regards to informed consent and IP considerations).

# History of changes

| Version | Date | Changes made | Author(s) |
|---|---|---|---|
| 0.1 | 09.03.2023 | FIRST DRAFT | Ilaria COLUSSI, Kurt MAJCEN (BBMRI-ERIC) |
| 0.2 | 16.03.2023 | Update after review by Robin Navest | Ilaria COLUSSI, Kurt MAJCEN (BBMRI-ERIC) |
| 0.3 | 17.03.2023 | Update after review by Laura Portell | Ilaria COLUSSI, Kurt MAJCEN (BBMRI-ERIC) |
| | | | |
| | | | |
| | | | |

**Statement of originality**

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# EXECUTIVE SUMMARY

The Data Management Plan (called "the project DMP" hereinafter) is a mandatory deliverable by the European Commission (EC) that has to be tailored to the specific project activities and which should document the relevant project repositories and how they are managed.

The Data Management Plan for EOSC4CANCER is developed by WP7 as an integral part of the governance structure, and also oversees the quality assurance of the project outcomes and alignment with FAIR data principles.

EOSC4CANCER will integrate and link existing resources referred to patient level data. Since EOSC4CANCER will not generate new data, the responsibility of data remains on local data access committees of each dataset that is linked for the project purposes.

This project DMP considers:

- the project data flow,
- how to comply with Open Science and FAIR data principles and
- how to integrate ethical and legal Requirements in the project.

This document is the EOSC4CANCER Data Management Plan (DMP) version 0.3 delivered in project month 06. This deliverable is based on the template and the guidelines provided by the European Commission for Horizon Europe. Such DMP is a living document that will be reviewed and updated periodically to ensure it remains up to date and thus serves the given purpose of documenting key elements in a research project's data lifecycle being comprised of collection, description, preservation, access and discovery of data.

The final version of the Project Data Management Plan will be provided as deliverable D7.4 in month 30 of the EOSC4CANCER project.

# 1 Introduction

The Data Management Plan (DMP) aims at showing how data are to be handled both during the research project, but also after the project has been completed.

The DMP considers the many aspects of data management, metadata generation, data preservation, and data analysis. It aims at informing about what happens during the data life-cycle as shown in **Figure 1**.

This DMP focuses on how data are produced by the project, how they are acquired, how they are processed after collection, what quality measures are adopted, how metadata are created or captured, how data are organized and annotated, how access and sharing occur, how data are stored, archived and preserved.

**Figure 1:** Data life-cycle

The DMP will be regularly updated throughout the lifetime of the project, and the latest updated and therefore currently valid version will always be available in the internal repository under the WP7 (Coordination, project management and ELSI framework) project folder.

# 2     Description of work accomplished

## 2.1     Project overview

EOSC4CANCER's mission is to connect cancer data (including cancer genomics, imaging, medical, clinical, environmental and socio-economic data) and make these data available for researchers and for different research projects in Europe.

Cancer has a complex nature that requires integration of advanced research data across national boundaries to enable progress. Therefore, access to data, knowledge and digital services – accessible across the European Research Area through federated infrastructures – is extremely important to achieve this mission.

For doing so, EOSC4CANCER makes use of existing federated and interoperable systems and intends to allow a secure identification, sharing, data processing and reusing of FAIR cancer data across borders.

The following three main stages in the individual patient journey of cancer patients constitute the object of attention for EOSC4CANCER, namely:

     i)        cancer causes and detection,
     ii)       diagnoses
     iii)      and treatment of (metastatic) cancer.

By tackling these three areas within one project, EOSC4CANCER's ambition is to move towards the integral approach to cancer that inspires the "Cancer Mission" in the context of EOSC. So, the idea is not to move on 'in separated silos' but to adopt a comprehensive, holistic view on cancer phenomena.

The project builds on and links to major national, European, and Global health and cancer initiatives that are at the basis of the European Cancer Mission. It connects distributed data sources by assembling a multidisciplinary consortium that builds on resources and expertise of the Life-sciences Research Infrastructures and cancer consortia, e.g. ICGC-Argo, GA4GH, 1+MG/B1MG, Cancer Core Europe, European Cancer Information System, European Network of Cancer Registries, Innovative Partnership for Action Against Cancer Joint Action and patients/survivors associations.

The project will use standards fully aligned with the European Open Science Cloud (EOSC[1]), such as "Guidance and policy on standards and tools to facilitate sharing and reuse of multimodal data (including imaging), cohort integration, and biosamples"[2] or "Report on data standards for observational and interventional studies, and interoperability between healthcare and research data"[3]. In particular, EOSC4CANCER fully adheres to EOSC philosophy

---

[1] EOSC-Portal, https://eosc-portal.eu/ (last seen: 28 February 2023)

[2] Boiten, Jan-Willem; Ohmann, Christian; Adeniran, Ayodeji; Canham, Steve; Cano Abadia, Monica; Chassang, Gauthier; Chiusano, Maria Luisa; David, Romain; Fratelli, Maddalena; Gribbon, Phil; Holub, Petr; Ludwig, Rebecca; Th. Mayrhofer, Michaela; Matei, Mihaela; Merchant, Arshiya; Panagiotopoulou, Maria; Pireddu, Luca; Richard, Audrey; Sanchez Pla, Alex; Schlünder, Irene; Tsamis, George; Wagener, Harald; Longo, Dario; Gorianin, Sergei. (2021). EOSC-Life Guidance and policy on standards and tools to facilitate sharing and reuse of multimodal data (including imaging), cohort integration, and biosamples. Zenodo. https://doi.org/10.5281/zenodo.4591011 (last seen: 28 February 2023).

[3] Canham, Steve, Ohmann, Christian, Boiten, Jan-Willem, Panagiotopoulou, Maria, Hughes, Nigel, David, Romain, Sanchez Pla, Alex, Maxwell, Lauren, Aerts, Jozef, Facile, Rhonda, Griffon, Nicolas, Saunders, Gary, van Bochove, Kees, & Ewbank, Jonathan. (2021). EOSC-Life Report on data standards for observational and interventional studies, and interoperability between healthcare and research data. Zenodo. https://doi.org/10.5281/zenodo.5810612 (last seen: 28 February 2023).

regarding open-sharing, standards, management of cancer data, data analysis, use and re-use of data for secondary use in research. Moreover, EOSC4CANCER will prepare EOSC services for cancer research and enrich EOSC with data, tools, and services from the cancer community.

It also aims to contribute to the creation of the European Health Data Space (EHDS), as provided by the EU proposal[4], and it links to other EU projects such as HealthyCloud[5], which builds the technical roadmap for European and national standards for access and sharing of health data for research (inputs from EOSC4CANCER to HealthyCloud will ensure long term sustainability of EOSC4CANCER); ELIXIR-CONVERGE[6], which helps standardise life science data management across Europe (EOSC4CANCER will build on its data management toolkit for scientists, extending it); CINECA[7], the Common Infrastructure for National Cohorts in Europe, Canada, Africa (since CINECA's experiences will be relevant for EOSC4CANCER); IMI FAIRPlus[8], which aims to develop tools and guidelines for making life sciences data FAIR, thus guiding the work in EOSC4CANCER when related to FAIRification of data; EJP RD[9], European Joint Programming for Rare Diseases (with which EOSC4CANCER will liaise, so to build compatible discovery and access solutions).

Finally, EOSC4CANCER will drive close alignment with technical and guideline development in EOSC-Life and will create direct synergies with the proposal for the cancer services related HORIZON-INFRA-SERV-2021-01-01 call.

As regards the project's methodology, EOSC4CANCER follows the patterns below:

i)      a selected number of use-cases will trigger specific questions associated with the three main stages of the Cancer Patient Journey;

ii)     these questions will be handled by some of the representative cancer analysis systems operating in the field;

iii)    these systems will request, via standard interfaces, the necessary data and methods;

iv)     the required data will be harmonised and made accessible under the proper legal conditions.

EOSC4CANCER is organized into three main pillars:

1. **TECHNICAL PILLAR:** WPs 1, 2 and 3 represent the technical pillar. WP1 focuses on federated data spaces, and so on secure data access across domains and jurisdictions. WP2 focuses on distributed harmonization and discoverability to

---

[4] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0197&qid=1669300754959 (last seen: 28 February 2023)

[5] HealthyCloud portal, https://healthycloud.eu/ (last seen: 16 February 2023). The HealthyCloud project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement Nº 965345.

[6] https://elixir-europe.org/about-us/how-funded/eu-projects/converge#:~:text=ELIXIR%2DCONVERGE%20is%20a%20project,management%20toolkit%20for%20life%20scientists (last seen: 28 February 2023)

[7] CINCECA Portal https://www.cineca-project.eu/, (last seen: 28 February 2023). CINECA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825775.

[8] FAIRPlus portal, https://fairplus-project.eu/ (last seen: 28 February 2023). FAIRPlus project has received funding from the Innovative Medicines Initiative Joint Undertaking under grant agreement No 802750. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA Companies.

[9] Portal EJP RD https://www.ejprarediseases.org/ (last seen: 28 February 2023). The EJP RD initiative has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°825575.

facilitate the interoperability. WP3 focuses on data analysis representing the software and computational stack.

2. **USE CASES PILLAR:** 5 different use-cases in WP4 constitute the second pillar. They drive and evaluate the solutions proposed by the technical pillar.

3. **COMMUNITY PILLAR:** it includes WP5, which ensures that the researchers are trained in the systems provided by the project, and WP6, which ensures the alignment with stakeholders and will produce a strategic roadmap for the long-term sustainability of the EOSC4CANCER platform.

# 3 Contribution towards project objectives

The project handbook[10] defines the project process that provides the framework to accomplish all projects objectives within the scope, budget and the required level of quality.

This deliverable has relevance for most project results and expected outcomes as listed below and the therefore indirectly also for the expected and wider impacts to a certain extent:

| Project results | Expected Outcomes (EO) | Expected Impacts (EI) | Wider Impacts |
|---|---|---|---|
| **#1** Access to major data resources, RI datasets and institutional repositories using open standardised methods, software and SOPs to make them available to the cBioPortal and MTBP visualisation and analysis systems as well as for other portals. | **EO3** At least 3 new FAIR and reusable datasets for each one of the EOSC4CANCER 5 uses cases along the patient journey | **EI3** Deployment of new protocols, implementation and SOPs to make accessible 8 main data types for cancer research. | Contribution to Europe's Beating Cancer Plan.

Contribution to attaining UN Sustainable Development Goal 3 on Good Health and Well-being.

Contribution to attaining UN Sustainable Development Goal 9 on Industry, Innovation and Infrastructure.

Contribution to implementing UNESCO Recommendations of Open Science.

Contribution to the European Strategy for Data.

Easier integration of methods in the analysis systems will improve the possibilities of Bioinformaticians, Computer Scientist and epidemiologist and other scientists to |
| **#2** Build the infrastructure for the identification, use and reuse of relevant cancer-related datasets while preserving data integrity, provenance and safety overcoming by design local legal, regulatory and technical requirements for data sharing, based on EOSC instruments. | **EO3** Demonstrated in at least 2 instances identification for each one of the 8 main data types necessary for the use cases, as well as with access to at least 3 transnational data instances for each one of the 5 EOSC4CANCER use cases | **EI3** New instruments, systems and protocols for future calls of the Cancer Mission. Uptake of the report on the general conclusions on the measures to identify and access transnational FAIR data sources relevant for cancer research overcoming data access restrictions. | |
| **#3** Harmonise and integrate data collections and systems, assuring their quality and provenance, incorporating the required big data and machine learning capacities | **EO1** Demonstrated access for use and reuse to at least 7 instances of data integration among the 8 main data types as required by the use cases | **EI2** New instruments, systems, and protocols for future calls of the Cancer Mission. A report on the software and protocols and recommendations for the harmonisation and integration of cancer data sets obtained from different technologies. | |
| **#4** Make accessible user environments and analysis portals enabling access to cancer data and analysis systems essential for the work in cancer across data types. | **EO1** Enhancing the capacity of existing cBioPortal and MTBP for cancer research and clinical trial handling for the use of the new data types (5 for cBioPortal and 2 for MTBP) | **EI2** Potential impact in basic cancer research and in future call of the cancer mission. Delivery of one protocol for each of the main data types and integrated data types to make accessible new datatypes via methods integrated in workflows for other portals developed by the community. | |
| **#5** Involve patients and cancer survivors' associations in the discussion together with the other stakeholders | **EO6** Build trust in the society on the capacity of transnational data access accelerate cancer research with direct | **EI2** Develop more than 30 instruments to be integrated in the EOSC ecosystem and sustained by the RIs and | |

---

[10] Villodres Raquel, Royo Romina (2022). EOSC4CANCER D7.1 - Project Management Handbook (v1.0), https://docs.google.com/document/d/1AED_LNC3-ZuQ7C4hldgyAVjbFeQoVHZn/edit (last seen 1 March 2023).

| #6 Provide additional connections in the Horizon Europe EOSC partnership and linking to the cancer research communities, | impact in cancer detection and treatment and seemly in decreasing cancer treatment inequalities. A final report to Cancer Core Europe, the Cancer Mission board and patient/survivor associations, as well as to professional societies and society in general, on experiences and success of EOSC4CANCER demonstrated by the practical results on the 5 use cases to improve the different steps along the cancer patient journey. | cancer research networks, that cover the different phases of distributed data finding, access, integration, and implementation in portals. Open the application to other tumour types, as well as to other chronic diseases and potential impact in other areas of biomedicine. | impact cancer research<br><br><br>Better clinical trial options for cancer patients<br><br><br><br>Facilities for organizing cohorts for cancer research and for starting new clinical trials |
|---|---|---|---|

**Table 1:** Summary of the Results, Expected Outcomes and Expected Impacts. For reference, please consult table in Grant Agreement – Annex 1 – DoA – Part B

The Grant Agreement of the EOSC4Cancer project defines several overall objectives which this DMP contributes to:

► **Objective 1:** Enable storage, sharing, access, analysis and processing of research data and other digital research objects from basic and clinical cancer research.

► **Objective 2:** Mobilise, interconnect and interoperate datasets relevant in cancer research.

► **Objective 3:** Make cancer research data and analysis systems easily accessible to basic and clinical scientists in the most used cancer analysis portals.

► **Objective 4:** Integrate digital tools, data analytics and Artificial Intelligence/Machine Learning tools for the analysis of cancer data in the cancer analysis portals.

► **Objective 5:** Contribute to the European Health Data Space (EHDS), the Horizon Europe European Open Science Cloud (EOSC) Partnership and the Cancer Mission.

# 4　Data Summary

EOSC4CANCER aims at systematizing efforts, linking the different data types, and building sustainable resources to facilitate equal access to data and methods.

EOSC4CANCER will work in a model of federated data access.

It will connect a set of interoperable nodes, e.g., European Cancer Centres, Research Infrastructures, Medical Centres, which provide access to FAIRified cancer-related data within a trusted users environment(s).

Data governance will build on the established institutional, local, national, and European frameworks. The selection of data and their connectivity will be driven by key use-cases in WP4 that demonstrate the value on all stages along the cancer patient journey and served in the visualization and analysis environments they are familiar with.

The overall aim is to coordinate data from established cancer networks and European research infrastructures, using open data standards (taking GA4GH and ICGC as references in the field of cancer), building on and further working from the EOSC ecosystem, and collaborating and heading towards the European Health Data Space and the European Cancer Mission.

During the project mainly four work packages (the technical WP1-3 and WP4 with the Use Cases) are involved in data flows as shown in **Figure 2**. The *WP5 Training* and *WP6 Community Engagement* are about accompanying measures and outreach beyond the project boundaries. As such they are not directly in touch with data, but a basic awareness of data flows and this DMP is helpful to them as well.
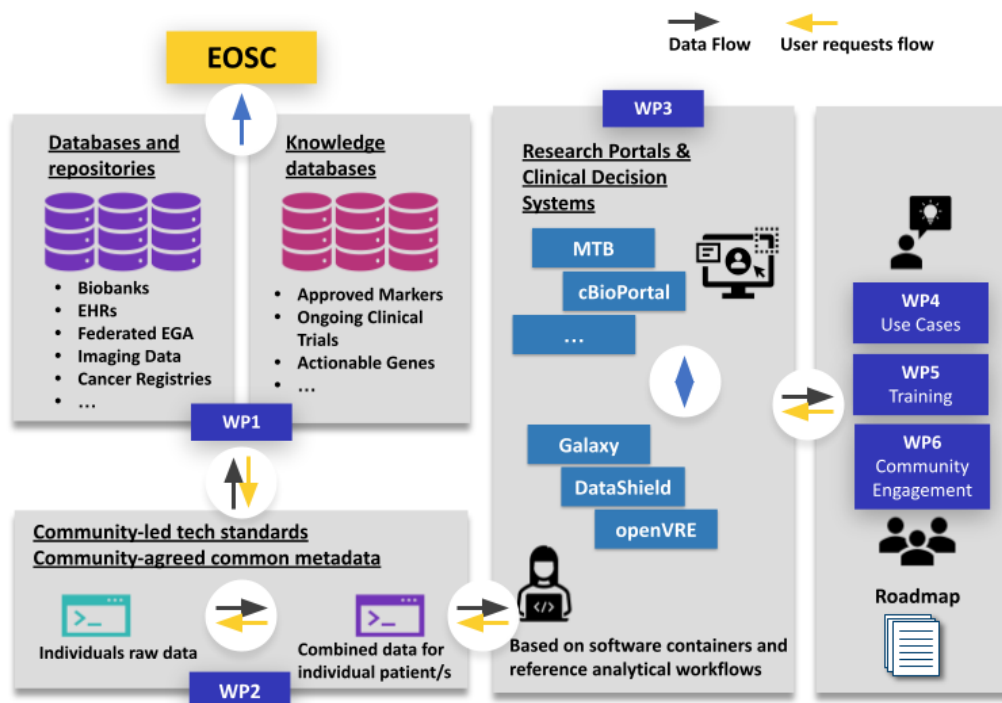


**Figure 2:** EOSC4Cancer overall concept

Therefore, the data flow is distributed among the different technical Work Packages (WP1-WP3) and WP4, flowing in a domain specific manner.

**WP1** will establish the fundamental components of a federated cancer data space. It is:

► developing a FAIR cataloguing platform for cohorts/biobanks/databanks/image repositories, their data dictionaries and formats/standards;

► connecting existing catalogues

► creating a living map of data access procedures onto the project website, including description of access interfaces, data use requirements and mapping resources FAIRness

► providing a set of cancer specific synthetic cohorts with multiple data modalities (molecular, imaging etc.) to facilitate testing and verification of federated clinical applications at scale.

► providing metadata descriptors to WP2 for data discovery (Beacon) and harmonization, and to WP3 for data access mechanisms

► upgrading the data access systems for participating resources in order to accelerate findability and accessibility for cancer multi-omics data.

The purpose is to build a federated approach where metadata is captured at source, annotated against standards, and integrated into a central search cache by using the MOLGENIS catalogue platform, in particular BBMRI Directory of Biobanks (also lead by BBMRI/UMCG), as well as by adapting best practice cataloguing systems emerging from H2020 projects for Cancer.

**WP3** will connect the questions from the use cases (WP4) and the data provision and interoperability (WP1 and WP2). The work will be based on existing computational methods served in the form of software containers interoperable in Virtual Research Environments (VRE) and workflow managers.

Therefore:

► a biomarker-driven clinical trials repository will be created. It will establish the procedures to ensure the handshake between the multidimensional biomarker-driven clinical trials information and clinical decision support systems

► a set of standardized, well-documented data upload pipelines and UIs for each of the data types as defined by the use-cases, and a SOP for secure data upload to and from distributed locations and/or from common European data repositories in connection with WPs 1 and 2, will be developed

► a reference installation of cBioPortal that can be used by use-cases (WP4) will be provided

► a packaged cBioPortal installation (e.g., in Docker), allowing oncology centres across Europe to implement a local cBioPortal installation, as well as demonstrators of its integration with novel data types to facilitate rapid translation of omics-driven and image-derived insights, will be distributed

► the dockerised version of the XNAT image processing tool to enable the consultation of images and radiomics results within the cBioPortal will be linked

► containerized software, e.g., using Docker and Singularity, and stable workflow descriptions, will be used for the reproducibility of results, and will facilitate access to

research software across different installations under the security protocols of the VREs.

**WP4** will provide cancer clinical research use cases driving data extraction, harmonisation, and analysis. The use cases will provide real life demonstrators of the benefits of transnational data infrastructure solutions for cancer research.

There are 5 use cases covering the Cancer Patient Journey. This WP acts as a driver triggering the technical work packages to provide the required data and services.

For doing so, WP4:

► identifies Cancer registries of particularly polluted areas that use 'geocoding' collection of data that can be used to examine spatial patterns of cancer incidence, as well as stage, survival and mortality, and to derive social status indicators (deprivation factors) and environmental characteristics

► delivers standardized and generalizable (i.e., beyond CRC) templates for the complex and longitudinal data handling in studies investigating localized cancer applicable in the framework of experimental precision oncology projects

► addresses the data infrastructures and format specifications required to analyse patients' tumour data by using a CDSS developed to inform about biomarker-driven precision medicine interventions

► builds on the Molecular Tumor Board Portal (MTBP), currently used in the Basket of Baskets trial (NCT03767075).

# 5      FAIR Data

Taking into account the above-mentioned project objectives (paragraph 3 of this DMP) and the work plan of EOSC4Cancer with its work packages (Paragraph 4 of this DMP) , we can observe that the following six components/steps of the project were deducted from the above objectives.

## 5.1      Making data findable

The first step is to obtain access to data belonging to major data resources, RI datasets and institutional repositories.

The ecosystem of cancer data is vast and not all data can be shared freely and immediately. So, it is important to consider different access conditions to controlled patient data, accreditation of data providers and existing national legislation and European regulation amongst many others.

Life Sciences Research Infrastructures (LS RIs) will act as proxies to gain access to different data types like screening data and other cancer repositories, as well as to connect with some databases and repositories, including the European Genome-phenome Archive (EGA), BBMRI-ERIC Federated Platform and Virtual Platform of European Joint Programme for Rare Disease (EJP RD).

EOSC4CANCER will build on top of these existing resources under the concept of "data federation".

## 5.2      Making data accessible: Infrastructure

The second step is to make data accessible and thus to build the infrastructure for the identification, use and reuse of relevant cancer-related datasets.

EOSC4CANCER will build a robust open infrastructure of data processing within the EOSC framework, enabling workflows to deal with complex distributed heterogeneous environments, integrate quality management and provenance standards, and provide reproducibility assurance.

The infrastructure for data analysis will be composed of Virtual Research Environments and workflow managers (primarily the Galaxy system but also adaptable to other management systems) based on software containers developed following ELIXIR Research Software Best Practices and the FAIR principles for research software [DOI:10.3233/DS-190026], acting as support and entry point to the analysis platforms.

Thus, EOSC4CANCER will fuse data from different sources and integrate different data types.

## 5.3      Making data interoperable: harmonisation and integration

The third component consists of making data interoperable though measures and standards of harmonisation and integration of data collections and systems.

EOSC4CANCER will establish guidelines and procedures for FAIR data management and access.

This will be done by incorporating and extending work carried on by global initiatives and within European projects, including the one resulting from the HORIZON-INFRA-SERV-2021-01-01

call, and the lessons learned from the European COVID-19 Data Platform and the recently funded EOSC BY-COVID project.

## 5.4    Increase data re-use: user environments and analysis portals

The fourth step is to make accessible user environments and analysis portals by enabling access to cancer data and analysis systems essential for the work in cancer.

EOSC4CANCER will deliver:

► Access to a broad range of cancer-related knowledge resources and catalogues across cancer domains integrated into the data analysis and visualisation platforms used by cancer research and clinical communities (i.e., cBioPortal for the analysis of primary tumours and Molecular Tumour Board Portal for metastatic cancer in the assignment to on-going clinical trials)

► An open and extensible metadata framework, adapted to EOSC use and the developments in LS RIs;

► A data and analysis framework for researchers, policymakers and patients/survivors that enables trust in results and credit to data submitters, workflow contributors and participant resources.

## 5.5    Stakeholder involvement

The fifth component consists of involving patients and cancer survivors' associations in the discussion together with the other stakeholders.

Engagement with stakeholders is, indeed, essential to incorporate FAIR and open data in cancer research, diagnosis and treatment.

Moreover, the demonstration of the practical utility of the EOSC4CANCER infrastructure and approach in clinical research use-cases will be proved by the associated development of standards and guidelines distributed to the cancer community, including patient associations, via outreach and training efforts.

## 5.6    Connecting to Horizon Europe EOSC partnership

The sixth component is to provide additional connections in the Horizon Europe EOSC partnership and linking to the cancer research communities. EOSC4CANCER will build on European data services, ESOC-Life and the EOSC open science practices, and will embed the research data management practice in national centres and European research infrastructures to ensure long-term sustainability of the platform. It will thus constitute the foundation for reproducible and trusted data analysis processes (such as federated learning) thus complying with the EU Cancer Mission.

# 6　　Methods: Open Science and FAIR sharing

The integrated ecosystem that EOSC4CANCER will deliver is based on established services and resources.

The project will not produce new patient level data but will work with sensitive data of cancer patients provided by the different contributing data sources, whose data access committees and processes remain in the end accountable for the basis for project operations, following a data federation model. In other words, data security remains the responsibility of the databases and repositories holding the source data (e.g., the EGA has a defined security process[11] and follows best practice guidelines aligned with the GA4GH Security Working Group, with the BBMRI-ERIC Policy for Access to and Sharing of Samples and Data[12] and with the OECD Council Recommendations on Health Data Governance[13]).

Moreover, the project will use synthetic non-identifiable data to facilitate the testing of the systems.

The EOSC4CANCER Ethics Advisory Board and the cancer patients and survivors' associations will be instrumental to guide the project in issues related with data use and privacy.

Data sharing and reuse will follow the FAIR principles[14] (**F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability).

As a final step, the project must ensure long-term sustainability, and this is possible through open-source platforms run on commonly used cloud-based infrastructures and connection of the components through open standards.


EOSC4CANCER shows a strong adherence to Open Science and FAIR principles, which are relevant for the project's long-term sustainability. Indeed, EOSC4CANCER engages with EOSC's common operational framework and ecosystem, implementing services envisioned in the EOSC Strategic Research and Innovation Agenda (SRIA)[15], and aligns with and builds on Open Science principles driving the implementation of SRIA.

Moreover, EOSC4CANCER's Project Partners have a track record in advocacy, policy and development of open access databases and repositories, see for example ECRIN metadata repository (MDR), ELIXIR Core Data Resources and Deposition Databases; FAIR standards and practices, e.g. FAIRsharing, RDMkit, FAIR Cookbook; Bioschemas, GA4GH; open software, technologies and transparent processing, e.g. WorkflowHub.eu, Galaxy project, RO-Crate; and open community-driven collaborative practices, e.g., ELIXIR Biohackathons, Galaxy smorgasbord. In addition, Project Partners are already leading global initiatives in FAIR Digital Objects (FDOForum, RDA Data fabric IG), FAIR Software (RDA FAIR4RS WG), and FAIR workflows (WorkflowsRI). EOSC4CANCER will contribute towards the FAIRification of cancer-related sources, building on our experience from previous projects, e.g. EOSC-Life, FAIRplus, HealthyCloud, CINECA.

---

[11] European Genome-phenome Archive: Security Overview https://ega-archive.org/files/European_Genome_phenome_Archive_Security_Overview.pdf (last seen: 16 February 2023)

[12] BBMRI-ERIC Policy for Access to and Sharing of Biological Samples and Data: https://www.bbmri-eric.eu/wp-content/uploads/AoM_10_8_Access-Policy_FINAL_EU.pdf (last seen: 16 February 2023)

[13] OECD Council Recommendations on Health Data Governance (last seen: 16 February 2023)

[14] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18 (last seen: 16 February 2023)

[15] European Commission, Directorate-General for Research and Innovation, Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC), Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2777/935288 (last seen: 28 February 2023)

Considering the concrete applications of Open Science and FAIR principles within EOSC4CANCER, it can be said that:

1. The following resources and open access tools will be used throughout the project for ensuring findability, discoverability and effective sharing:
   a. open-source research software connected with computational capabilities (e.g., Galaxy, openVRE)
   b. other systems like Nextflow, or
   c. extensible open-source imaging informatics software platform dedicated to imaging-based research (i.e. installation of XNAT to be used in the use cases).

2. Open data standards (taking GA4GH and ICGC as references in the field of cancer) will be adopted too. Indeed, open standards for cataloguing and indexing will allow access to an open and extensible metadata framework, adapted to EOSC use and the Life Sciences Research Infrastructures (LS RIs) developments.

3. Through the offered training activities, EOSC4CANCER will be fully aligned to the EOSC skills and roles as defined by the Digital skills for FAIR and Open Science Report (DOI:10.2777/59065)[16], adapted to fit the Cancer Data stakeholders. EOSC practises (e.g., EOSC Interoperability framework[17], EOSC Enhance D4.3 "Analysis of existing research data cataloguing efforts towards integrated discovery"[18]) will also be followed.

---

[16] https://www.eoscsecretariat.eu/sites/default/files/files/eosc_users.png (last seen: 28 February 2023)

[17] EOSC-IF: https://eosc-portal.eu/eosc-interoperability-framework (last seen: 28 February 2023)

[18] Carole Goble, & Nick Juty. (2021). Analysis of existing research data cataloguing efforts towards integrated discovery. Zenodo. https://doi.org/10.5281/zenodo.4693217 (last seen: 28 February 2023)

# 7 Conclusions

Data management provides help to minimize potential errors that can appear with introduced processes and established policies, be it for projects or within organizations. It is important for creating trust to what happens with data. As such a well-established and accurate Data Management Plan sets the ground for successful collaboration between the WPs across the EOSC4CANCER project.

This first version of the Data Management Plan gives an initial description of the data flows within the EOSC4CANCER project. It shows that EOSC4CANCER works in a model of federated data access, where different data types are linked and sustainable resources to facilitate equal access to data and methods aim to be build.

Data access committees and processes for each dataset remain, in any case, accountable for the basis for project operations, as provided by a data federation model. Thus, data security remains the responsibility of the databases and of the repositories holding data.

# 8    Next steps

This version of the DMP, a living document, will be regularly updated taking into account the ongoing work and the submitted deliverables and reached milestones in the project. The current DMP will thus be reviewed to ensure it continues to be fit for purpose and that any changes introduced regarding the repositories and the processes to handle data are incorporated into this document.

The final update of the Data Management Plan (as deliverable D7.4) is scheduled for month 30 of the project as defined in the EOSC4CANCER Grant Agreement.