This partnership has received funding from the European Union's "EURATOM" research and innovation program under the 101061037 grant agreement

PIANOFORTE Partnership

European Partnership for Radiation Protection Research

Horizon-Euratom – 101061037

WP9 Open Call 1
Project DISCOVER-D1

Lead Author: Simonetta Pazzaglia (ENEA)

With contributions from: Emiliano Fratini (ENEA), Simone Moertl (BfS), Katalin Lumniczky (NNGYK)

Reviewer(s):

Liz Ainsbury (UKHSA, WP5 Lead)

PIANOFORTE Coordination Team (IRSN, WP1 & WP9 lead)

PIANOFORTE Coordination Team (IRSN,	**WP 9**	**T 9.1**
Project Acronym	**DISCOVER**	
Deliverable nature:	**Report**	
Dissemination level: (Confidentiality)	**Public**	
Contractual delivery date:	**Month 6 – 30 July 2024**	
Actual delivery date:	**Month 6 – 30 July 2024**	
Version:	**1**	
Total number of pages:	**11**	
Keywords:	**Data management plan, DISCOVER**	
Approved by the PF coordinator:	**PIANOFORTE Month 26**	
Available on PF website	**July 2024**	

Abstract

PIANOFORTE DISCOVER Data Management Plan (DMP) describes the main strategies and techniques for data sharing according to the FAIR (findable, accessible, interoperable and re-usable) principles. It also includes security, regulatory and ethical issues of data management.

The deliverable describes the types of data generated as well as operational procedures concerning how data generated or used in DISCOVER will be handled, stored, shared; it also describes the process of quality assurance.

PIANOFORTE DISCOVER DMP was created on the template of the DMP of PIANOFORTE Partnership and considers all its basic principles and recommendations, thus it forms an integral part of the PIANOFORTE Partnership main DMP.

Content

1. Aim of the document

1.1 Introduction

PIANOFORTE-DISCOVER was selected through the PIANOFORTE first open call answering the topic "Developing a knowledge base for a better understanding of disease pathogenesis of ionising radiation-induced cancer to improve risk assessment". The full title of the project is: "Dissecting radIation effectS into the Cerebellum micrOenVironmEnt driving tumour pRomotion". Its main purpose is to study the impact of radiation-induced changes in the microenvironment and the influence of related cell communication processes on carcinogenesis. The main objectives of DISCOVER are the following:

- To study how different cerebellar populations, such as granule cell precursors, the Medulloblastoma (MB) cell of origin, and astrocytes, microglia and endothelium, representing microenvironmental components, respond to moderate (2 Gy) and low (0.1 Gy) radiation doses and contribute to tumour formation.
- To evaluate the effect of the microenvironment in transmitting radiation signals driving carcinogenesis.
- To conduct a comprehensive analysis of various types of data, including morphology, function, tumourigenesis and omics data.
- To investigate secretome, as well as extracellular vesicles from exposed tissue and their specific bioactive cargo for their role in mediating radiation tumourigenesis
- To integrate analysis of DISCOVER animal data and publicly available human brain cancer data aims to identify patterns/signatures for Medulloblastoma (MB) development.

WP (Task)	Description	DM responsible
WP1 (Task 1.1-1.2)	**Investigating changes in the cerebellum microenvironment following cranial irradiation**	ENEA, Emiliano Fratini
WP2 (Task 2.1-2.2)	**Dissecting radiation effects in MB target cells and ME in cerebellum in vitro**	NNGYK, Katalin Lumniczky
WP3 (Task 3.1-3.2)	**Dissecting the impact of radiation effects on GCPs and their ME ex -vivo**	ENEA, Emiliano Fratini
WP4 (Task 4.1-4.4)	**Role of secreted factors and EVs in radiation -induced signaling**	BfS, Simone Moertl
WP5 (Task 5.1-5.3)	**Functional tests of EV and other relevant pathways**	NNGYK, Katalin Lumniczky

PIANOFORTE-DISCOVER Data Management Plan (DMP) describes the principles for the management of data within the project and was created in full compliance with the policy adopted by PIANOFORTE using the PIANOFORTE DMP (as described in Deliverable D5.2) as a template.

1.2 Objectives

PIANOFORTE-DISCOVER Data Management Plan (DMP) defines the strategies and tools for managing the data generated during and after the end of the project. The DMP clarifies the technical and organisational aspects of managing project data and deliverables. It is a descriptive document, which defines the main strategies and tools to ensure secure, transparent, and efficient storage, maintenance, sharing, and use of data. It was constructed in alignment with the project objectives and data structure. It represents a living document, which will be updated during the project, if significant changes occur.

1.3 Purpose and scope of this deliverable

The purpose of this deliverable is to define a frame for optimal data management to allow making project data findable, accessible, interoperable, and reusable (FAIR).

2. Data summary

2.1 The purpose of the data collection/generation

The data collected in DISCOVER will allow dissemination, reuse and accountability of the research performed within the project with the aim to support the realisation of the main objectives of the PIANOFORTE partnership including the DMP in Task 5.5 (https://pianoforte-partnership.eu/workpackages/wp5-infrastructures-and-data-management-for-radiation-protection-research) in relation to FAIR and open science.

2.2 The relation of DISCOVER to the objectives of the PIANOFORTE partnership

DISCOVER is related to the following specific objectives of PIANOFORTE:

- Specific objective 1: “To improve scientific understanding of the variability in individual radiation response and health risk of exposure”
- Specific objective 2: “To support regulations and implementation of the BSS and improve practices in the domain of low dose exposures of humans and the environment by better understanding and reducing uncertainties in risk estimates”

2.3 Types and formats of data generated/collected

Types of data generated within DISCOVER: experimental, observational and numerical data in the form of radiobiological, omics, epidemiological and imaging data sets. Data will be spreadsheets such as excel files, word documents, PDFs, images (TIFF, JPG), tab delimited text files (TSV) and RAW file format (.raw) for Omics analysis.

2.4 Re-use of existing data

Most of the data collected within DISCOVER will be newly generated data.

Within WP1 and WP4 of the project we will perform bioinformatic analysis to identify affected pathways, protein-miRNA-RNA interaction, principal Biological Processes and molecular functions involved. We will use: GO databases for gene ontology enrichment analysis; STRING, PANTHER, REACTOME, g:profiler, Enrichr and miRbase Databases for network analyses; Single Cell Expression Atlas and Panglao DB Databases for cell type identification on the bases of transcriptome profiles and cell type markers.

In addition, within WP5 we will use also already available datasets deposited in publicly available databases (Gene Expression Omnibus, GEO; The Cancer Genome Atlas, TCGA; National Cancer Institute Genomic Data Commons, GDC; Pediatric Brain Tumor Atlas, PBTA; Children's Brain Tumor Tissue Consortium, CBTTC) in order to match the crucial events, and key cellular and soluble components, necessary to promote MB development in murine models with brain tumour data in humans.

2.5 Origin of data

Data within DISCOVER will be generated from the following sources:

- Experimental data generated by Omics (Single cell gene expression profiling, Proteomics, Genome Methylation), tumorigenesis and radiation-induced signalling analysis (neuroinflammation, angiogenesis, extracellular vesicles) *in vivo* within WP1. Generated data will refer to the transcriptome profile of different cell types within the cerebellum, global proteome profiles, DNA methylation patterns of functional loci, microscopic characteristics, gene expression profile, signalling and extracellular vesicle characteristics and radiosensitivity.
- Experimental data generated from 2D and 3D *in vitro* cell models of cerebellum co-cultures studying the interaction of MB cell precursors (Granule Cell Precursors, GCPs) with their environment within WP2. Generated data will include proliferation, DNA-damage, apoptosis, senescence, permeability, tube formation and stem cell analysis using enzymatic activity assays, immunocytochemistry staining, qPCR gene expression analysis.
- Experimental data from *ex-vivo* 3D models of Organotypic Cerebellar Slice Culture (OCSC) within WP3. Data will be generated using enzymatic activity assays, immunohistochemistry staining, qPCR gene expression analysis.
- Experimental data from secretome, Extracellular Vesicle (EV) and EV cargo characterization within WP4. This data will include Omics data (Proteome, miRNome) and multiplex immunoassay for secreted factors.
- Experimental data on the role of EVs in transmitting radiation signalling *in vivo* will be generated and validated within WP5. Data will be obtained by Western blot analysis and immunohistochemistry.
- Data generated by the re-use and re-analysis of datasets from publicly available databases as described in subchapter 2.4 (WP1, WP4 and WP5).

2.6 Expected size of the data

DISCOVER will generate 1) big data from the multiomic analysis of irradiated cerebellum (WP1) (scRNA-seq data 500Mb to 1 Gb in .tar format, 10 GB proteome), proteome (30 GB)

and miRNome (5-10Mb) analysis for the characterization of EV cargo and multiplex immunoassays for secretome characterisation (2MB) (WP4); 2) small datasets related to all the other analysis (microscopic analysis, qPCR, enzymatic activity assays, western blot analysis, etc.- Mbytes).

2.7 Data utility: to whom it will be useful

The data generated within DISCOVER will be useful to radiation safety regulators and investigators concerned with the biological and health effects of radiation in environmental, medical and industrial contexts. The data will contribute to the aims of several EU platforms including the MEENAS platforms (MELODI, ALLIANCE, EURADOS, SHARE, EURAMED, NERIS), International Organisations, such as the International Atomic Energy Agency, International Agency for Cancer Research and the International Commission on Radiological Protection, as well as national regulators and agencies.

3. FAIR

DISCOVER will respect the "FAIR" principles for data handling so that data are findable, accessible, interoperable, and reusable to facilitate effective open science and sharing among researchers, stakeholders, and policymakers.

All publications will be full open access, which will facilitate access to the raw data, either within the publications or with a link to a permanent record, e.g. a Digital Object Identifier (DOI). Early open sharing of research data will be realised through presentation at international conferences. Research data will be made available to the community when the associated manuscripts are published.

3.1 Making data findable, including provisions for metadata

In order for the data to be findable, DISCOVER will provide well-structured data information. Data produced and/or used in the project will be identifiable and findable through metadata, using the DOI as a unique identifier. Naming conventions, search keywords and version numbering will be clear and well-documented, and the process of data generation will be well explained.

3.1.1 Outlining the discoverability of data

Data generated in DISCOVER will be deposited by default in the STORE database (https://www.storedb.org/store_v3/), unless scientific journals where project results are intended to be published require other databases). STORE is maintained by the German Federal Office for Radiation Protection (BfS). It is a stable and accessible platform for archiving and sharing of data outputs from radiation research. The well-established database already contains a range of relevant data and it is well-used by the radiobiology community.

ORCID (Open Researcher and Contributor ID) identifier will be used to identify the individuals who deposited the data. Data referenced in publications and reports will be linked via DOI

numbers. Documentation on the source of the data and tools needed to read or use the data will be provided as free text narrative in the database.

3.1.2 Outlining the identifiability of data and reference to the standard identification mechanism

A DOI and a database-specific identifier will be used to identify the data deposited in the database.

3.1.3 Outlining naming conventions used

The Technical Information Library Services (TILS) naming convention will be used to deposit records in the database. This will include the number of the work packages and tasks/subtasks from which they were generated and version number using the ISO8601 data format standard.

3.1.4 Outlining the approach towards search keywords

Data will be annotated with standard ontologies and datatypes as detailed above.

3.1.5 Outlining the approach for clear versioning

Versions will be specified in the naming convention and attached as metadata.

3.1.6 Standards for metadata creation

Descriptive metadata standards will be used to ensure consistency in describing and managing data.

3.2 Making data openly accessible

Most data generated by DISCOVER will be freely available in the public domain. Data not freely available will be shared through a request to the principal investigator, after ascertaining that the legal and ethical conditions for data sharing are fulfilled.

3.2.1 Specifying which data will be made openly available

As stated above, most data generated within DISCOVER will be made publicly available. Those data which pose conflicting intellectual property and privacy issues (datasets owned by commercial entities, personal or sensitive health-related data, or local legal restrictions do not allow open sharing) will not be publicly shared.

3.2.2 Specifying how the data will be made available

Data generated within DISCOVER will be made publicly available through the following means:

- Full, open-access publications which include the raw data, either within the publications or with a link to a permanent record, e.g. a DOI. Research data will be made available to the community when the associated manuscripts are published.
- Early open sharing of research data will be realised through presentation at international conferences.

3.2.3 Specifying what methods or software tools are needed to access the data

Whenever possible, data will be made available in formats that require commonly used software such as text (.docx, .txt, .tsv, .csv), spreadsheet (.xlsx, .tsv, .csv), image (.jpg, .tiff) readers, and

Adobe Acrobat (.pdf). If certain data require dedicated software that are not freely available, they will be made available upon request from the data generators.

3.2.4 Specifying where the data and associated metadata, documentation and code are deposited

Data generated within DISCOVER will be deposited in:

- The STORE database. The project data will be deposited in STORE as data files or as references to accession numbers/DOIs for the data in other databases, specified in the publications. Each dataset and data item will be assigned a persistent STORE ID and a DOI which can be used for reference.

3.2.5 Specifying how access to data will be provided

The restrictions to the data will be imposed for:

- Embargo until publication
- Reasonable embargo until intellectual property issues are determined and resolved as specified in the PIANOFORTE guidelines and the individual contracts issues under the Open Calls.
- Proprietary data will be accessed by negotiation with the data holders and owners and by the establishment of licensing agreements
- Personal and genomic data will be restricted due to General Data Protection Regulation (GDPR), privacy legislation and ethical considerations. Data may be accessible by negotiation with the data holder and relevant Institutional data access and ethics committees.

3.3 Making data interoperable

3.3.1 Assessing the interoperability of data

Data formats will follow community norms and will be compliant with generally available software.

3.3.2 Specifying the use of standard vocabulary for all data types present in DISCOVER data set, to allow inter-disciplinary interoperability

DISCOVER will use standard vocabulary from Open Biological and Biomedical Ontology (OBO) for all data types.

3.4 Making data re-useable

3.4.1 Specifying how the data will be licenced to permit the widest possible reuse

Most data generated in DISCOVER will be released under the CC-BY licence, allowing anyone to copy, distribute, transmit and adapt work and make commercial use of work under the condition that the user must attribute the work in the manner specified by the author or licensor.

3.4.2 Specifying when the data are made available for re-use

In general, DISCOVER data will require an embargo period until publication. In case-by-case situations data might be made available to specific collaborators subject to individual agreements.

3.4.3 Specifying whether the data produced and/or used in the project is useable by interested parties, in particular after the end of the project

Most of the data generated by the project will be re-usable by interested parties after the embargo period specified above. These data will be available through the STORE database for a period of minimum ten years. Human health-related and personal data will not be re-usable by interested parties.

3.4.4 Data quality assurance

Responsibility for data quality rests with the coordinator of DISCOVER, Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), who appoints Emiliano Fratini (emiliano.fratini@enea.it) as data manager. The data manager will oversee, and check files uploaded to chosen platforms when they are archived for sharing. Before each periodic report, the data manager will review data shared during the reporting period as well as file availability and integrity.

3.4.5 Specifying the length of time for which the data will remain reusable

The data will be available through the STORE database for a period of minimum ten years.

4. Allocation of resources

4.1 Estimating the costs for making data FAIR and describing the potential source to cover the cost

Costs related to data management within DISCOVER are limited and will be covered by the budget of the project partner who is the data generator, as part of the overheads claimed by the respective partner. Individual investigators within the projects have personal responsibility for the curation of their shared data.

4.2 Identifying responsibilities for data management in DISCOVER

As already specified, the responsibility for data management rests with the Project Coordinator of DISCOVER, ENEA. The main data manager from ENEA will be Emiliano Fratini (emiliano.fratini@enea.it), who will work in collaboration with data managers appointed by each scientific partner within DISCOVER (BfS and NNGYK). Data managers of project partners will be responsible for managing data generated by the respective partner or generated in collaboration with other partners, but coordinated by the respective partner, while data manager of ENEA will in addition also be responsible for data quality assurance as described above.

4.3 The costs and potential value of long-term preservation

Since data generated by DISCOVER are intended to be stored preferentially on the STORE database or other databases maintained without charge for the user, the costs of long-term preservation are expected to be minimal.

The main value of the long-term storage of the data relies in the fact that DISCOVER will generate primary Omics data related to radiation response in the cerebellum and tumorigenesis, which will be suitable for re-analysis in the future.

5. Data security

As mentioned above, data generated within DISCOVER will be stored in the STORE database or other similar international repository databases that all have state-of-the-art data backup and recovery systems and data are mirrored at other geographically distant locations and are frequently backed up.

Part of the data will, in addition, be stored on institutional databases. Each partner institution has its own policy for data management and security. Local servers are used for regular data backups, from where data are moved to either public repositories or to institutional platforms for long-term storage using standard backup systems.

Personal data and sensitive health data will be stored in compliance with the relevant national regulations also in conformity with EU-based ISO standards. Such data will not be kept on unencrypted workstations, laptops or other media.

6. Ethical and legal issues

Research within DISCOVER is regulated by both international and national legal and ethical rules. All partners will conform to the International, European and National legislations in all aspects of the research and ensure that legal and ethical requirements will be fulfilled. The ethical standards of guidelines of Horizon Europe as described in the European Code of Conduct for Research Integrity (ALLEA, 2017) will be rigorously applied, regardless of the country in which the research is carried out.

WP1, WP2, WP3, WP5 analyses require the use of animals (as well as primary cells or *ex-vivo* tissues derived from animals) to fulfil several tasks in DISCOVER project. For this purpose, mice will be irradiated in ENEA and tissues (cerebella, blood) will be sampled. Moreover, non-irradiated mice will be used to collect tissue (cerebella) for cell isolation and OCSC establishment. Procedures on animals will be performed only after obtaining the approval of the ENEA Animal Welfare Body (AWB) and authorisation from the Italian Ministry of Health and will be conducted according to the requirements of the Italian legislations related to animal welfare. The legislations are EU-conform and are governed by Directive 2010/63/EU on the protection of animals used for scientific purposes. In short, handling of experimental mice will be performed respecting the three major ethical principles of work with experimental animals, the 3Rs: reduction (the number of experimental mice will be kept at the minimum and unnecessary repetitions of experiments will be avoided), refinement (procedures will be optimised to cause no or minimal pain for the animals) and replacement (wherever possible, *in vitro* co-culture and *ex-vivo* experiments will be performed in order to avoid use of live animals). Considering a moderate severity for the procedures, data related to animals (i.e.

number of animals used) will be submitted to Italian Ministry of Health for retrospective analysis, at the end of the procedures.

DISCOVER is aware of further relevant guidance, codes and regulations, including:

- Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes;

- ARRIVE guidelines, a set of recommendations to ensure the full and transparent reporting of research involving animals;

- The Nuremberg Code (1946) addressing volunteer consent and proper acting;

- The Revised Declaration of Helsinki in its last version of 2013;

- The convention for the protection of human rights and dignity of human being with regard to the application of biology and medicine called the "Convention on Human Rights and Biomedicine" (Council of Europe, 1997) and the additional protocol on the prohibition of cloning human beings (1998); and its additional protocol on biomedical research (2005):

- Recommendation CM/Rec (2016)6 on research on biological materials of human origin was adopted by the Committee of Ministers of the Council of Europe:

- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.:

- Code of practice on secondary use of medical data in scientific research projects, 2014;

- UN Convention on the Rights of the Child (1990);

- Opinions of the European Group on Ethics in Science and New Technologies (as from 1998). https://ec.europa.eu/research/ege/index.cfm?pg=reports;

- EU Guidelines for Ethics in Social Sciences and Humanities (http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020_ethics-soc-science-humanities_en.pdf);

- The New Brunswick Declaration: A Declaration on Research Ethics, Integrity and Governance resulting from the 1st Ethics Rupture Summit, Fredericton, New Brunswick, Canada (2013);

- The Respect Code focused in socio-economic research Ethics Guidelines for Trustworthy AI: HLEG on Artificial Intelligence. European Commission, Brussels. (https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai)