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 IMAGEOMICS-D 1 – Data Management Plan (DMP) -

Lead Author: Katalin Lumniczky (NNGYK)
With contributions from: Géza Sáfrány (NNGYK), Melanie Fachet (OVGU), Claudia Scotti and Giorgio Baiocco (UNIPavia)

Reviewer(s):
Liz Ainsbury (UKHSA, WP5 Lead)
PIANOFORTE Coordination Team (IRSN, WP1 & WP9 lead)

Work package / Task	**IMAGEOMICS WP 5**	**T5.3**
Deliverable nature:	**Data Management Plan, Report**	
Dissemination level: (Confidentiality)	**Public**	
Contractual delivery date:	**Month M3 - 30 April 2024**	
Actual delivery date:	**Month M4 – 08 July, 2024**	
Version:	**2**	
Total number of pages:	**13**	
Keywords:	**Data management plan, IMAGEOMICS**	
Approved by the PF coordinator:	**PIANOFORTE Month 25 – July 2024**	
Available on the PF website	**July 2024**	

Disclaimer:

The information and views set out in this report are those of the author(s). The European Commission may not be held responsible for the use that may be made of the information contained therein.

Abstract

PIANOFORTE IMAGEOMICS 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 IMAGEOMICS will be handled, stored, shared; it also describes the process of quality assurance.

PIANOFORTE IMAGEOMICS DMP was created on the template of the DMP of PIANOFORTE Partnership and takes into account 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-IMAGEOMICS was selected through the PIANOFORTE first open call answering the topic "Individualised diagnostic and therapeutic procedures with regard to optimisation of the benefit/risk ratio". The full title of the project is: "Optimizing Benefit/Risk Ratio in Breast Cancer Diagnosis and Radiotherapy: Identifying Molecular, Cellular and Imaging Signatures of Breast Cancer Heterogeneity to Improve Personalized Therapeutic Strategies for Synergistic Treatment Combinations". The Project partners are the following: National Center for Public Health and Pharmacy (NCPHP), Budapest, Hungary; EUROPEAN ALLIANCE FOR MEDICAL RADIATION PROTECTION RESEARCH (EURAMED), Wien, Austria; Otto von Guericke University (OvGU), Magdeburg, Germany; CENTRO DE INVESTIGACIONES ENERGETICAS, MEDIOAMBIENTALES Y TECNOLOGICAS (CIEMAT), Madrid, Spain; Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain; Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA), Rome, Italy; University of Pavia (UNIPV), Pavia, Italy. The project is coordinated by NCPHP. NCPHP, EURAMED, CIEMAT and ENEA are POMs, while OVGU, IsGlobal and UNIPV serve as associated entities (AE) in the PIANOFORTE project.

The main purpose of IMAGEOMICS is to improve radiological protection of breast-cancer (BC) patients and thus to provide solutions for a future optimisation of the European radiation protection Basic Safety Standards. The main objectives of IMAGEOMICS are the following:

- To investigate how radiotherapy (RT) influences immunogenic heterogeneity of BC cells of different molecular subtypes using in vitro and in vivo approaches.
- To test the applicability of nanoparticles for X ray fluorescence computed tomography (XFCT) to be used for the detection of BC heterogeneity.
- To identify local and systemic signatures that predict patient benefit from combined RT and immunotherapy and test their clinical applicability.
- To integrate data retrieved from experimental models and human studies with existing epidemiological data to build up a protocol for optimal patient stratification.

To achieve these aims, the work is performed in five Work Packages: WP1 - Investigating the effect of radiotherapy on the phenotype, gene expression signatures, and secretory profile of BC cell lines in vitro (Lead: NCPHP); WP2 - Investigating the impact of RT on the interaction between BC cells, immune cells and other components of the tumor stroma using in vitro and in vivo mimicking approaches (Lead: OvGU; WP3 - Screening for molecular signatures of RT application and immune system activation in human BC samples (Lead: UniPV); WP4 - Clinical applicability and long-term follow-up studies of patients – proof of concept (Lead: ISGlobal); WP5 - Coordination, dissemination and E&T strategy, data management plan (Lead: NCPHP).

PIANOFORTE-IMAGEOMICS 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-IMAGEOMICS 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 organizational 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 course of 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 IMAGEOMICS 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.

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

IMAGEOMICS is related to the following specific objectives of PIANOFORTE:

- Specific objective 1: “To innovate in ionising radiation based medical applications combating cancer and other diseases by new and optimised diagnostic and therapeutic approaches improving patient health and safety and supporting transfer of the R&I outcome to practice.”
- Specific objective 2: “To improve scientific understanding of the variability in individual radiation response and health risk of exposure.”

It is also linked to both the EU SAMIRA action plan, Europe’s Beating Cancer Plan and Horizon Europe Mission on Cancer, since it helps patients in benefiting from increased RT treatment efficacy and diminished risk of side effects, and it improves the availability of effective treatments for patients based on multi-modal synergistic anti-cancer treatment solutions.

2.3 Types and formats of data generated/collected

Types of data generated within IMAGEOMICS: experimental, observational and numerical data in the form of radiobiological, clinical, epidemiological and imaging data sets. Data will be spreadsheets such as excel files, word documents, PDFs, images (TIFF, JPG), gene expression and RNA raw sequence files, mass spectrometry files.

2.4 Re-use of existing data

The majority of the data collected within IMAGEOMICS will be newly generated data.

Within WP4 of the project we will use also already available datasets deposited in publicly available databases. A systematic review of all available databases will be conducted to generate a set of data on breast tumour types, tumour characteristics, treatment and outcome of treatment. The open access databases will be identified through clinical trials registries (such as cliniclatrials.gov, cancer research UK and EORTC) and other database registries (such as database commons, cbioportal) as well as further databases identified by project partners.

In addition, detailed search on different histological, cellular, molecular and imaging biomarker categories will be performed using literature search and search of data deposited in relevant databases. The aim of this data retrieval is to identify biomarkers with predictive value for combined RT and immunotherapy approaches in breast cancer. This approach will help in the identification of those cellular and molecular tumour characteristics at individual patient level that predict a favourable response to combined treatment protocols involving radiotherapy and immunotherapy and thus contributes to an improved patient stratification.

2.5 Origin of data

Data within IMAGEOMICS will be generated from the following sources:

- Experimental data generated by 2D, 3D and organ-on-a-chip in vitro culture of commercially available breast cancer cell lines within WP1 and WP2. Generated data will refer to the microscopic and immunological characteristics, cell vitality and morphology, gene expression profile, signalling characteristics, radiosensitivity, interaction with their environment and capacity to uptake various nanoparticles and microvesicles.
- Experimental data including X-ray based imaging data generated in WP2 using laboratory mice and murine breast cancer models. Generated data will refer to the immunological characteristics and radiosensitivity of the murine tumour models.
- Histological and molecular characteristics of tumour samples collected from patients, and associated clinical records of possible previously administered therapies (WP3)
- Experimental ex-vivo/in-vitro data generated in WP3 concerning the response to radiation and immune interaction with blood cells of cancer cells from patients' samples
- Data generated by the re-use and re-analysis of datasets from publicly available databases as described in subchapter 2.4 (WP4).

2.6 Expected size of the data

IMAGEOMICS will generate 1) small datasets related to the immunological parameters, radiation response and nanoparticle uptake characteristics of breast cancer cell lines and 2) big data from the multiomic characterisation of breast cancer samples from patients.

2.7 Data utility: to whom it will be useful

The data generated within IMAGEOMICS 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 Data

IMAGEOMICS 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, IMAGEOMICS 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 IMAGEOMICS will be deposited by default in the STORE database (unless scientific journals in which scientific results generated within the project are intended to be published require other databases). STORE is hosted by the German Federal Office for Radiation Protection (BfS). It is a central access portal to information from radiobiology research distributed across scientific institutions, a well-established database already containing 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 ISO8601data 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 IMAGEOMICS 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 IMAGEOMICS 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 IMAGEOMICS 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 readers, Excel, and Adobe Acrobat. 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 IMAGEOMICS 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. 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 IMAGEOMICS data set, to allow inter-disciplinary interoperability

IMAGEOMICS 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 IMAGEOMICS 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, IMAGEOMICS 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 any 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 third parties.

3.4.4 Data quality assurance

Responsibility for data quality rests with the coordinator of IMAGEOMICS, National Center for Public Health and Pharmacy (NNGYK), who appoints Balázs Gyebrovszki (gyebrovszky.balazs@nngyk.gov.hu) 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 IMAGEOMICS 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 IMAGEOMICS

As already specified, the responsibility for data management rests with the Project Coordinator of IMAGEOMICS, NNGYK. The main data manager from NNGYK will be Balázs Gyebrovszki (gyebrovszky.balazs@nngyk.gov.hu), who will work in collaboration with data managers appointed by each scientific partner within IMAGEOMICS (OVGU, UNIPavia and ISGlobal). 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 NNGYK 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 IMAGEOMICS 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 IMAGEOMICS will generate primary data related to molecular, cellular and imaging heterogeneity of breast cancer and related radiation response, which will be suitable for re-analysis in the future.

5. Data security

As mentioned above, data generated within IMAGEOMICS 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 IMAGEOMICS 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.

WP3 analyses require the use of human breast cancer samples (as well as primary cells derived from these samples) to search for molecular signatures of radiotherapy application and immune system activation. For this purpose, fresh breast cancer tissue samples will be collected from surgical specimens (breast resections or mastectomies for breast cancer) available at the Anatomic Pathology Unit of the University Hospital Policlinico San Matteo. When possible, blood samples from patients who underwent surgical procedures will be obtained as well, for isolation of the peripheral lymphocyte pool. Tumour sampling and peripheral blood collection will be performed only after obtaining the patients informed consent and only if the tumor size and specimen volume are sufficient to perform investigations needed to achieve a breast cancer diagnosis and an adequate biologic characterization of the tumour. All the clinical, pathological and biological data collected for the study will be anonymised. The protocol will be submitted to the ethical committee of Fondazione IRCCS Policlinico San Matteo Pavia.

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

- 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 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)

Experiments and sample collection from mice in WP2 will be performed after approval by the corresponding national body from Hungary and will be conducted according to the requirements of the Hungarian 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 3D experiments will be performed in order to avoid use of live animals).