HORIZON-HLTH-2023-CARE-04-03
Environmentally sustainable and climate neutral health and care systems

NetZeroAICT

Digital Contrast for Computerised Tomography
-Towards Climate Neutral and Sustainable Health Systems-

Starting date of the project: 01/12/2023
Duration: 48 months

= Deliverable D2.1 =

Active data management plan produced (v1)

Due date of deliverable: 29/02/2024
Actual submission date: 29/02/2024
Requested revision submission date: 24/11/2025

Work Package: 2
Responsible Work Package Leader: Anders Nordell, CMRAD
Responsible Task Leader: Jonatan Eriksson, CMRAD
Current task leader: Carlos Santin, CMRAD
Version: V1.1

Dissemination level		
PU	Public – fully open (automatically posted online on the Project Results platforms)	X
SEN	Sensitive — limited under the conditions of the Grant Agreement	
Classified R-UE/EU-R	EU RESTRICTED under the Commission Decision No2015/444	
Classified C-UE/EU-C	EU CONFIDENTIAL under the Commission Decision No2015/444	
Classified S-UE/EU-S	EU SECRET under the Commission Decision No2015/444	

Co-funded by the European Union

AUTHOR

Author	Institution	Contact (e-mail, phone)
Carlos Santín	CMRAD	carlos@cmrad.com
Anders Nordell	CMRAD	anders@cmrad.com
Ernest Casany Pujol	CMRAD	ernest@cmrad.com

DOCUMENT CONTROL

Document version	Date	Change
V0.1	29.02.2024	First draft
V1.0	29.02.2024	Edits by Scientific Coordinator
V1.1	31.10.2025	Revision as per PO and reviewer comments on de-identification processing and residual risk, legal and harmonization aspects

VALIDATION PROCESS

Reviewers		Validation date
Work Package Leader	Anders Nordell	29.02.2024 New version 19/11/2025
Project Manager	Martina Nešverová/ Mariana Pacheco Blanco*	29.02.2024 New version 21/11/2025
Scientific Coordinator	Regent Lee	29.02.2024 New version 21/11/2025
Coordinator	Anders Nordell	29.02.2024 New version 19/11/2025

***current Project manager**

DOCUMENT DATA

Keywords	Data management
Point of Contact	Name: Carlos Santín Partner: CMRAD Phone: +34 661 931 386 E-mail: carlos@cmrad.com
Delivery date	29/02/2025 Revised version 19/11/2025

DISTRIBUTION LIST

Date	Version	Recipients
29.02.2024	V1.0	EC, all partners via Google Drive
19.11.2025	V1.1	All partners
24.11.2025	V1.1	EC

DISCLAIMER

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

This document is the first version of the NetZeroAICT Data Management Plan (DMP) for Work Package 2 (WP2). It sets out how data will be collected, Pseudonymized, transferred, curated, accessed, stored, shared and preserved across the Project, in line with Horizon Europe requirements and the FAIR principles (Findable, Accessible, Interoperable and Reusable).

NetZeroAICT builds on the existing AICT Consortium. The Clinical Partners contributing data to NetZeroAICT are already part of the broader AICT initiative and have signed a Joint Controller Agreement (JCA) governing their joint responsibilities under the General Data Protection Regulation (GDPR). The AICT Data Protection Impact Assessment (DPIA) and Legitimate Interest Assessment (LIA) provide a pre-existing risk and legal basis analysis which NetZeroAICT inherits and further specifies for the purposes of this Project.

For operational reasons, and to enable large-scale, lawful and efficient ingestion of data from multiple Clinical Partners, the Project uses a two-layered governance approach that combines:

- Local Controller–Processor relationships between each Clinical Partner and Collective Minds Radiology AB (CMRAD), which operates the technical infrastructure and Repository; and
- Joint controllership among the NetZeroAICT Partners for the secondary use of Pseudonymized Data for research and AI development, as governed by the AICT JCA, extended to NetZeroAICT.

At each Clinical Partner site, CM-Connect is installed as a local edge component. It acts as a "proxy" within the hospital's infrastructure, ensuring that direct identifiers never leave the local environment and that only minimized, Pseudonymized datasets are transmitted to the NetZeroAICT Repository. The Clinical Partner remains the only party able to link Pseudonymized Data back to identifiable patients, while CMRAD operates the Repository as a Processor, under detailed Data Processing Agreements (DPAs). Once data are available in the central Repository, the NetZeroAICT Partners (including AI development partners) act as Joint Controllers for the scientific processing performed within the Project.

The DMP explains how this architecture reduces re-identification and reconstruction risk while fully acknowledging that Pseudonymized Data remains Personal Data and that the rights and freedoms of Data Subjects must be respected. The document clarifies:

- The legal bases for secondary use for scientific research at Controller level, aligned with GDPR Articles 6 and 9 and applicable local/national laws.
- The allocation of roles and responsibilities among Controllers, Joint Controllers and Processors.
- The mechanism for exercising Data Subject rights in a setting where only the Clinical Partner holds the re-identification key.
- The planned harmonization of imaging metadata and any associated clinical variables, using international standards and controlled vocabularies.
- How FAIR principles will be implemented in a way that is compatible with the nature of CT imaging data and the residual re-identification risk.

The Repository will primarily contain Pseudonymized CT imaging data in DICOM format, together with structured metadata needed for cohort definition, AI training and validation. Given the inherent identifiability of high-resolution medical images, and the risks of singling out or linkage with external datasets, NetZeroAICT does not treat the imaging data as anonymized. Instead, the Project uses strong Pseudonymization, contextual controls and a closed, secure processing environment operated by CMRAD to keep the residual risk at an acceptable level for scientific research. Open access to raw imaging data is therefore not envisaged. However, the Project will identify which aggregated or derived outputs (for

example, summary statistics, model performance metrics or, where appropriate, suitably risk-assessed derived datasets) can be shared more broadly, while keeping data protection risks within acceptable limits.

This first version of the DMP is delivered in month 3 of the Project and will be updated at least at month 24, and further if needed, as new datasets, processing activities and risk assessments become available. It should be read in conjunction with D1.1 “Legal framework signed with 10 Clinical Partners” and D8.11 “Adaptations to legal frameworks or contracts – report v1” (delivered at M12), which provide more detail on the contractual and data protection framework underpinning NetZeroAICT.

List of Abbreviations and Acronyms

CE = Contrast Enhanced

CT = Computerized Tomography

CTA = CT Angiogram

DICOM = Digital Imaging and Communications in Medicine

DMP = Data Management Plan

FAIR = Findability, Accessibility, Interoperability, and Reusability (FAIR).

IV = Intravenous

PACS = picture archiving and communication system

RCM = Radiocontrast Media

AICT = Artificial Intelligence for Computerized Tomography (parent consortium)

AWS = Amazon Web Services

DPA = Data Processing Agreement

DPIA = Data Protection Impact Assessment

EHDS = European Health Data Space

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

JCA = Joint Controller Agreement

LIA = Legitimate Interest Assessment

MSA = Master Service Agreement

WP = Work Package

Table of Contents

DEFINITIONS **6**

1. INTRODUCTION **7**

2. DATA SUMMARY **9**

2.1. PURPOSE OF THE DATA COLLECTION/GENERATION AND ITS RELATION TO THE OBJECTIVES OF THE PROJECT 10
2.2 TYPES AND FORMATS OF DATA THE PROJECT WILL COLLECT AND GENERATE 11
2.3 DATA PROTECTION 12
2.5. WHAT IS THE ORIGIN OF THE DATA? 20
2.5. WHAT IS THE EXPECTED SIZE OF THE DATA? 21
2.6. TO WHOM MIGHT IT BE USEFUL ('DATA UTILITY')? 21

3. FAIR DATA **21**

3.1. OVERVIEW 21
3.2 DATA DESCRIPTION 21
3.3 ENSURING FAIR DATA PRINCIPLES 22

4. ALLOCATION OF RESOURCES **25**

5. DATA SECURITY, PERSONAL DATA BREACH, AND INCIDENT MANAGEMENT **26**

5.1. WHAT PROVISIONS ARE IN PLACE FOR DATA SECURITY (INCLUDING DATA RECOVERY AS WELL AS SECURE STORAGE AND TRANSFER OF SENSITIVE DATA)? 26
5.2. IS THE DATA SAFELY STORED IN CERTIFIED REPOSITORIES FOR LONG TERM PRESERVATION AND CURATION? 26
5.3 PERSONAL DATA BREACH RESPONSE AND INCIDENT MANAGEMENT 26

6. ETHICAL ASPECTS **28**

7. CONCLUSIONS **29**

8. DEGREE OF PROGRESS **29**

9. DISSEMINATION LEVEL **29**

Definitions

For the purposes of this DMP, the following capitalized terms have the meanings set out below. Where these terms are already defined in the Grant Agreement or the Consortium Agreement, the definitions in those instruments prevail in case of inconsistency.

Project means the NetZeroAICT Project "Digital Contrast for Computerized Tomography – Towards Climate Neutral and Sustainable Health Systems".

Consortium means the NetZeroAICT Consortium as defined in the Grant Agreement and Consortium Agreement.

Clinical Partner means a Beneficiary or Associated Partner that operates a healthcare institution or imaging provider and contributes clinical CT data to the Project.

AI Development Partner means a Beneficiary or Associated Partner whose primary role is the development, training, validation or evaluation of AI models within the Project.

Platform means the cloud-based technical environment operated by CMRAD that hosts the NetZeroAICT Repository and provides tools for storage, curation, cohort building, AI training and validation.

Repository means the central storage environment within the Platform in which Pseudonymized Data and associated metadata from the Clinical Partners are stored and made available for processing within the NetZeroAICT Work Packages.

Personal Data has the meaning given in Article 4(1) GDPR.

Special Categories of Personal Data means Personal Data referred to in Article 9(1) GDPR, including health data and imaging data processed in NetZeroAICT.

Controller (or Data Controller) has the meaning given in Article 4(7) GDPR. In this Project, each Clinical Partner is a Controller for the clinical data it collects in its routine care and for the secondary use of those data, in line with its national implementation of the GDPR and applicable ethics approvals.

Joint Controller has the meaning given in Article 26 GDPR. In this Project, the NetZeroAICT partners, who jointly determine the purposes and essential means of processing Pseudonymized Data in the Repository, act as Joint Controllers under the Joint Controller Agreement originally concluded in AICT and extended to NetZeroAICT.

Processor has the meaning given in Article 4(8) GDPR. CMRAD acts as a Processor for each Clinical Partner under a DPA for the technical and organizational operations necessary to ingest, Pseudonymize, store and provide controlled access to data in the Repository.

Pseudonymization and **Pseudonymized Data** have the meanings specified in Article 4(5) of the GDPR. Pseudonymized Data are Personal Data that can no longer be attributed to a specific Data Subject without additional information, provided that such additional information is kept separately and protected by appropriate technical and organizational measures. In NetZeroAICT, the mapping between hospital identifiers and NetZeroAICT subject identifiers is held only by the Clinical Partner.

Data Subject means an identified or identifiable natural person whose Personal Data are processed in the Project.

Data Protection Legislation means the GDPR and, as applicable, the UK GDPR, national implementing laws and any other local data protection frameworks that apply to a given Controller or processing context.

Data Management Plan (DMP) means this document, which describes the policies, procedures and technical and organizational measures for data management in the Project, and which will be updated during the Project's lifetime.

Joint Controller Agreement (JCA) means the agreement between the AICT (and extended NetZeroAICT) partners that allocates responsibilities and describes Joint Controllership for the secondary use of Pseudonymized health data within the Platform.

Data Processing Agreement (DPA) means an agreement concluded under Article 28 GDPR between a Controller and a Processor, describing the subject matter, duration, nature and purposes of processing, the type of Personal Data and categories of Data Subjects, and the obligations and rights of the Controller.

Data Protection Impact Assessment (DPIA) means an assessment required under Article 35 GDPR for processing that is likely to result in a high risk to the rights and freedoms of natural persons.

Legitimate Interest Assessment (LIA) means an assessment of the legitimate interests pursued by a Controller and the balancing of these interests against the interests or fundamental rights and freedoms of the Data Subjects, where Article 6(1)(f) GDPR is relied on as a legal basis.

CM-Connect means the local technical component deployed at the Clinical Partner's premises, acting as a "proxy" between local PACS or hospital systems and the Platform. It performs local Pseudonymization and data minimization before transfer to the Repository.

1. Introduction

Deliverable D2.1 "Active data management plan produced (v1)" is part of Task 2.1 "Data management plan" in Work Package 2. It describes how NetZeroAICT will manage the large-scale CT imaging data needed to develop and validate AI models capable of synthesizing "CT Digital Contrast" and reducing the environmental footprint of radiological contrast use and CT imaging.

In line with Article 17 and Annex 5 of the Horizon Europe Grant Agreement, this DMP explains:

- Which types of data will be collected, re-used, processed and/or generated
- Which methods, standards and governance structures will be applied
- How data will be handled during and after the Project, including retention and long-term preservation
- Under which conditions data or derived outputs may be shared or made available ("as open as possible, as closed as necessary")
- How data protection, ethical and security requirements will be taken into account.

NetZeroAICT is closely linked to the AICT Consortium, which established the initial legal and data protection framework for multi-site CT data sharing. The AICT DPIA and LIA, together with the AICT Joint Controller

Agreement, form the foundation upon which NetZeroAICT builds. However, practical experience in AICT showed that initiating large-scale data transfer across sites required a more tailored operational and contractual mechanism, particularly to increase local assurance around accountability and security during the ingestion phase. NetZeroAICT therefore introduces a Controller–Processor framework between each Clinical Partner and CMRAD, combined with the pre-existing Joint Controller structure for secondary research use in the Platform.

This DMP is conceived as a living document. The present version is produced in month 3 and will be updated at least once (D2.2) in month 24, and further if required, to reflect new datasets, risk assessments, standards and governance decisions taken by the Consortium.

2. Data Summary

NetZeroAICT aims to reduce the environmental footprint created by Computerized Tomography (CT), a common diagnostic imaging modality, by developing trustworthy AI capable of generating "CT Digital Contrast" from non-contrast enhanced images. CT is one of the most widely used diagnostic imaging techniques in healthcare and is estimated to result in several megatons of CO_2 emissions globally. Iodinated RCM used for CE CT scans also contributes significantly to pharmaceutical waste. NetZeroAICT aims to reduce this environmental footprint by enabling safe reduction in RCM use across Europe by 2033.

The Project will curate a large Repository of CT imaging data (including both CE and non-CE CT studies) contributed by Clinical Partners from the AICT Consortium. The Repository is expected to contain data from up to 1,000,000 patients, with approximately half of the studies being CE and half non-CE, across several body regions (head/neck, chest, abdomen, pelvis and limbs). These data will support data curation (WP3), AI model development (WP4) and validation (WP5), as well as lifecycle monitoring and environmental analyses in other Work Packages

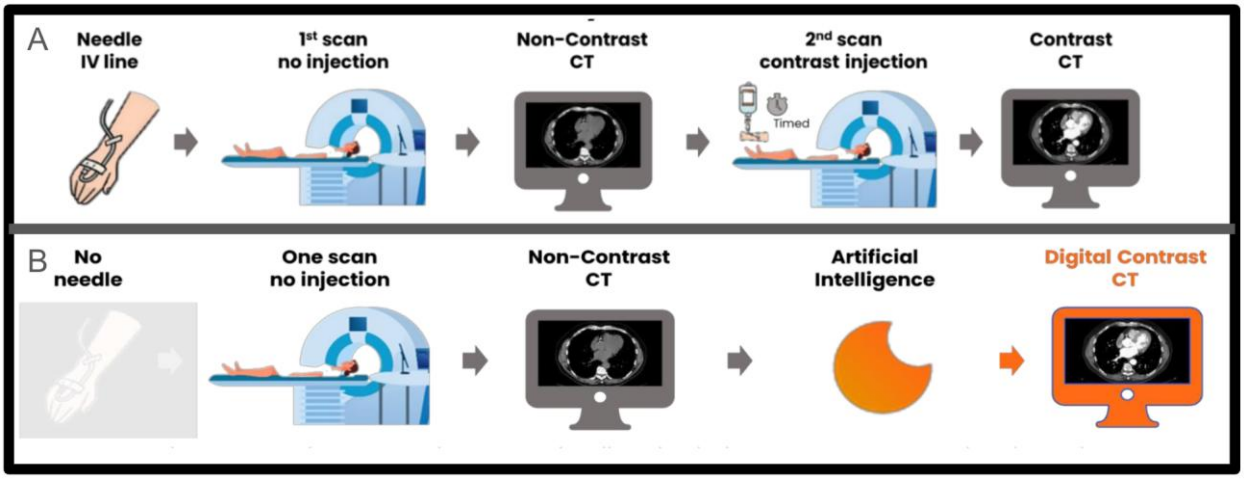

Figure 1: A) Conventional Computerized Tomography (CT) workflow. A needle and intravenous (IV) line is set on the patient, a first scan without the administration of contrast agent, rendering a Non-contrast enhanced CT, for the second scan contrast agent is injected prior to the scan rendering a contrast enhanced (CE) CT scan. B) In the workflow proposed there is no needle and IV line, and no second scan. An AI network is instead used to render the Digital Contrast CT.

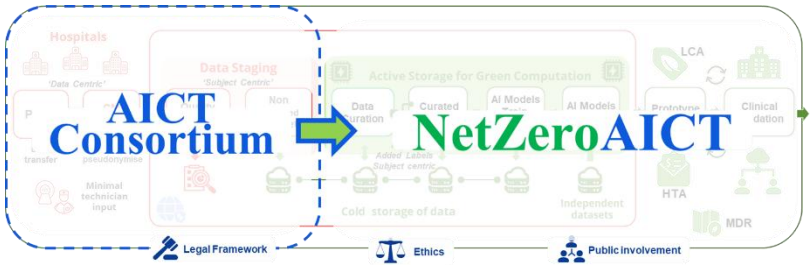

Figure 2: Transition from AICT towards NetZeroAICT Consortium.

The Project has nine objectives, separated into a number of deliverables that will be delivered from seven work packages (WPs). The purpose of this document is to outline the complete set of data-related activities of the NetZeroAICT Project. The document will be updated regularly based on modifications and new details related to the NetZeroAICT data.

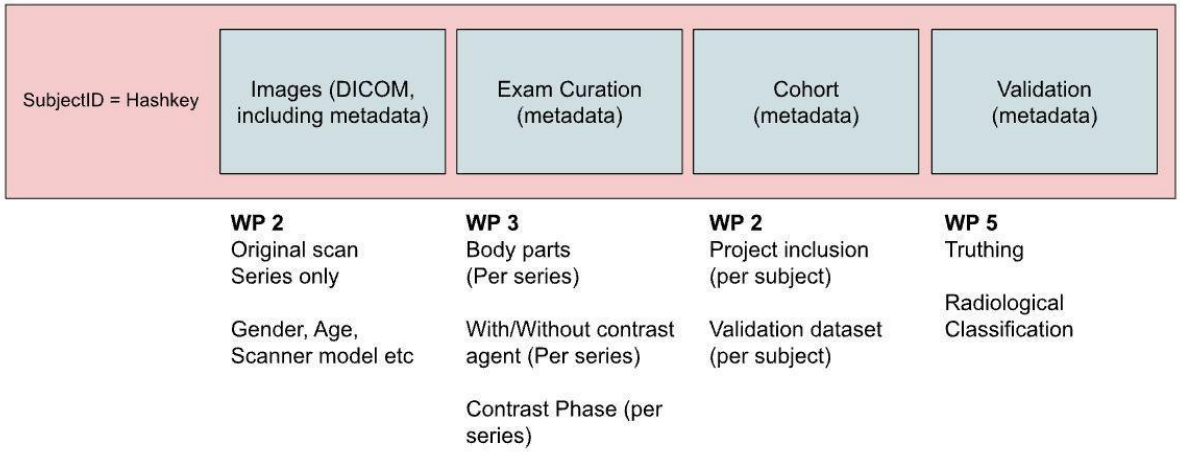

Figure 3: Schematic diagram of the image and metadata in three of the different work packages (WP).

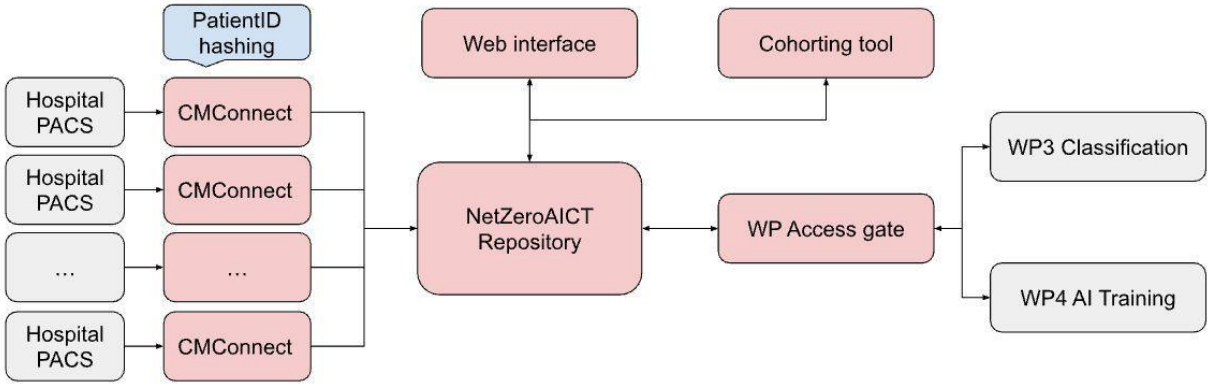

Figure 4: Schematic diagram of the data flow within the Project. Each Clinical Partner sends their image data using the CM-Connect where the pseudonymization is being performed, to the Project data stage, in which the different work package projects may interact with the data.

2.1. Purpose of the data collection/generation and its relation to the objectives of the Project

The main input data consist of clinical CT imaging acquired by Clinical Partners as part of routine care for diagnosis, treatment planning or follow-up. For inclusion in the Project, these studies are selected in a pathology-agnostic manner to create a diverse and representative dataset in terms of patient demographics, underlying conditions, scanner vendors and models, acquisition parameters and contrast usage patterns.

A sufficiently large and heterogeneous cohort is needed to:

- Train AI networks that can learn the mapping between non-CE and CE CT (Digital Contrast) across multiple anatomical regions and scanners
- Validate performance for safety and robustness, including across subgroups (for example, different age groups, sexes, comorbidities or regions)
- Support environmental analyses and health-system modelling related to RCM use reduction.

In WP2, the Repository is set up and populated. WP3 refines and enriches the data, including automated and manual classification to support cohort selection. WP4 and WP5 use the curated datasets to train and

validate AI models. The initial ingestion milestones (D2.3, D2.4, D2.5) are crucial for achieving the scale required for the scientific and environmental objectives.

2.2 Types and formats of data the Project will collect and generate

In the NetZeroAICT project, imaging data will be collected, transferred and stored according to the **D**igital **I**maging and **Co**mmunications in **M**edicine or DICOM standard[1]. Images stored in the DICOM format, hold both the pixel values that constitute the actual image and a vast amount of metadata ("DICOM tags"), along with unique IDs (UIDs). Each Patient, Study, Series and Instance hold a unique identifier. Figure 5 outlines the hierarchy of the DICOM standard.

The hierarchy of the DICOM standard is PATIENT, STUDY, SERIES and IMAGE (or INSTANCE)

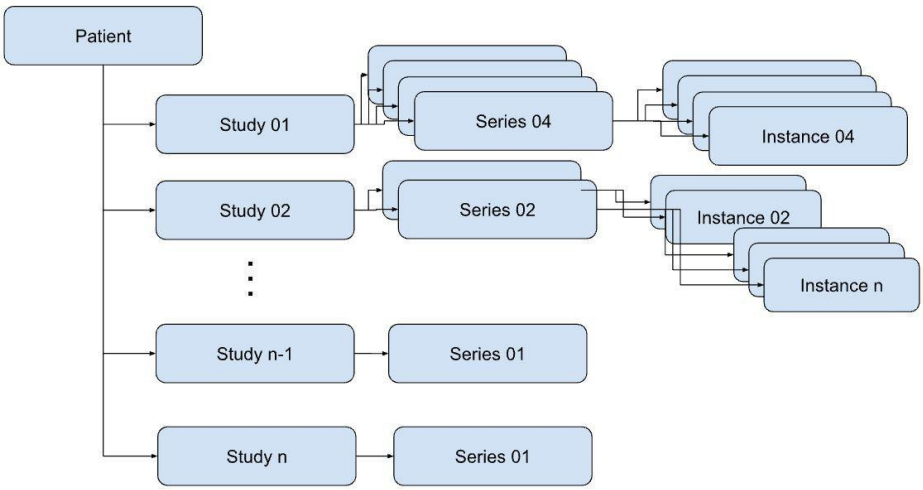

Figure 5: The hierarchy of DICOM starts with a patient, each patient can have one or several studies. Each study may consist of one or more series and each series contains one or more instances or images.

The Project will primarily collect, transfer and store imaging data in DICOM format. DICOM objects contain both pixel data and a structured header that encodes acquisition parameters, scanner information, and certain clinical and demographic data. Patient-level, study-level, series-level and instance-level identifiers are included in the original clinical archive but are transformed as part of the Pseudonymization process. The Project distinguishes two broad categories of data within the Repository:

1. **Image Data**:
   CT image series in DICOM format, including non-CE and CE scans across the supported anatomical regions. The pixel data are preserved with diagnostic quality to support AI training and validation.

2. **Associated Data and Metadata**:
   Structured information necessary for cohort definition, AI model development and evaluation, and environmental analyses. This includes, for example:
   - technical acquisition parameters (scanner vendor, model, protocol, slice thickness, tube current, kVp, reconstruction kernel, contrast phase)
   - high-level anatomical labels and series descriptors
   - RCM usage information (for example, whether RCM was used, basic dosage categories if available)

[1] https://www.dicomstandard.org/

- limited, pre-defined clinical variables where needed to support specific analyses (for example, indication categories, basic outcome measures)
- internal Repository identifiers (Site ID, Subject ID, Study ID, Series ID) required for traceability and cohort building.

Non-essential identifiers and free-text fields that carry a high re-identification risk but limited scientific utility (for example, detailed dates and times, free-text comments) are either not ingested or are transformed during pseudonymization (for example, date shifting or generalization).

Imaging data are stored using the DICOM standard; associated data and metadata are stored within the Platform's database structures, which will be aligned with controlled vocabularies and standard terminologies where feasible (see section 3 on FAIR).

2.3 Data protection

The NetZeroAICT Project involves handling a large amount of health-related data (as defined in article 4(15) of the GDPR), which are considered special categories of data (as per article 9 of the GDPR) and is subject to multiple restrictions and requires careful consideration of the associated risks to protect the privacy rights and freedoms of the Data Subjects and patients. Furthermore, the Project aims to process data from various sources and jurisdictions, which require compliance not only with the GDPR but also with other local regulations, such as Australia's Privacy Act 1988, Brazil's LGPD, or the UK's UK GDPR.

For the purpose of archiving the above, a system is created through that installation of a local piece of software (CM-Connect), that acts as a proxy that ensures that the local identifiable personal data remains at the local infrastructure (e.g. hospital) and all data is Pseudonymized (de-identified) and minimized before being transferred to the central NetZeroAICT Repository, ensuring compliance with the local laws and requirements and ensuring data protection and privacy risks minimization in the NetZeroAICT Repository (depicted in Figure 6 below). A detailed description of the data privacy and IT security assessment is provided in Deliverable D1.1.

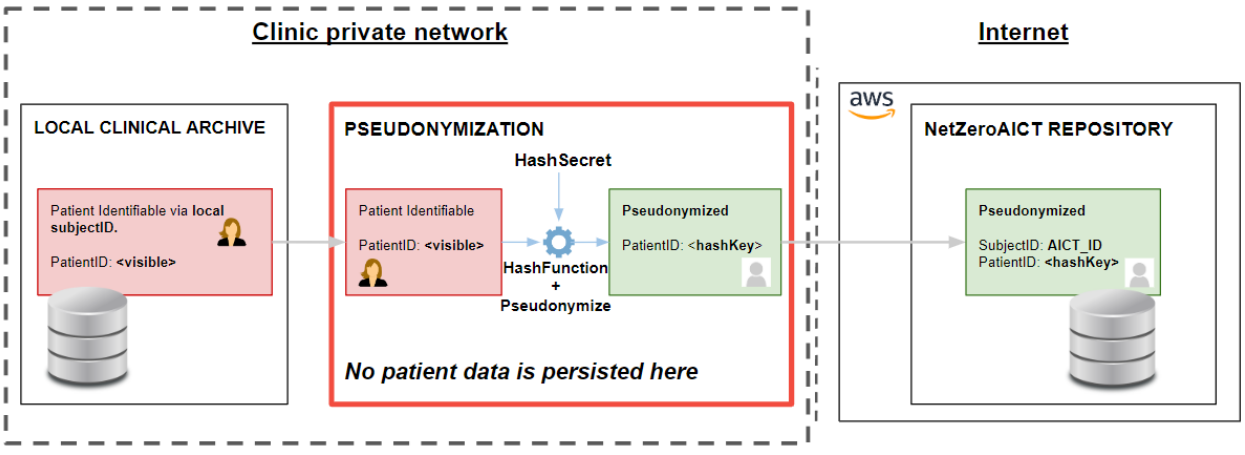

Figure 6: Schematic diagram of data flow with pseudonymization from Clinical Partner to NetZeroAICT Repository.

2.3.1 Nature of the data and applicable legislation

The NetZeroAICT Project processes Special Categories of Personal Data in the sense of Article 9(1) GDPR, namely health data and CT images that relate to identifiable individuals when considered at source. The Consortium therefore applies the strictest applicable standards, using the GDPR as the baseline and taking into account any local laws and guidance in the jurisdictions of the Clinical Partners, including non-EU partners, where transfers to the EU Repository must also comply with relevant international data transfer requirements (e.g., LGPD is Brazil's General Data Protection Law and Australian Privacy Act 1988, Australian Privacy Principles, etc.).

2.3.2 Roles and responsibilities (Controllers, Joint Controllers and Processor)

At each Clinical Partner site:

- The Clinical Partner is the Controller for the Personal Data it collects and stores in its local systems (for example, PACS and electronic medical records)
- CMRAD acts as a Processor for the purpose of receiving, Pseudonymizing, storing and technically managing Pseudonymized Data in the Repository on behalf of the Clinical Partner, under a DPA.

Once the Pseudonymized Data is available in the Repository for use within NetZeroAICT:

- The NetZeroAICT Partners that jointly decide on the purposes and essential means of secondary processing (for example, which cohorts to build, which AI models to train, how to validate and interpret results) act as Joint Controllers under the AICT JCA, extended to NetZeroAICT. This includes both Clinical Partners and AI Development Partners, in accordance with the Consortium Agreement and JCA.
- CMRAD continues to act as Processor for the Joint Controllers when operating the Platform and implementing their documented instructions.

This layered model reflects the practical reality that data flow must be enabled at scale via a robust Controller–Processor framework, while the scientific processing and decision-making remain under Joint Controllership of the NetZeroAICT Partners.

2.3.3 Pseudonymization process and re-identification risk

To reduce re-identification risk while preserving scientific utility, NetZeroAICT applies strong pseudonymization at the Clinical Partner premises before any data leaves the local environment.

The process can be summarized as follows:

- CT data are retrieved from the local PACS or equivalent system and passed to CM-Connect, which is installed within the Clinical Partner’s infrastructure and under its control.
- Direct identifiers (for example, patient name, hospital number, address, contact details) are removed from DICOM headers. An internal Subject ID used in NetZeroAICT is generated, typically by applying a cryptographic hash to a local identifier combined with a site-specific secret or salt.
- The mapping between local identifiers and Project Subject IDs are stored securely within the Clinical Partner’s systems and is not shared with CMRAD or other Partners.

- Additional minimization and risk-based transformations are applied to indirect identifiers where necessary. This may include, for example, replacing dates with relative measures (age at examination, or year and month instead of exact dates), generalizing certain geographic or institutional codes, and excluding or recoding rare combinations of fields that could increase singling-out risk.

- The Pseudonymized DICOM series and associated metadata are transmitted over encrypted channels to the Repository hosted in the EU.

The Project explicitly recognizes that CT imaging data, even after Pseudonymization, remain Personal Data and that a non-negligible risk of re-identification or "reconstruction" persists. Reconstruction risk in this context covers both:

- the possibility of identifying individuals from the rich anatomical detail present in CT images (for example, unique skeletal features, implants, or facial structures in head/neck scans) or by linkage with external datasets; and

- model-based risks such as membership inference or model inversion attacks against trained AI models.

Because of this, NetZeroAICT does not claim that the Repository is anonymized. Instead, it relies on a combined approach:

- strong Pseudonymization and minimization at source

- a closed, secure processing environment with strict access controls, logging and governance

- risk-aware design of derived datasets and outputs

- a DPIA and LIA at AICT level that will be further detailed and, where necessary, updated for NetZeroAICT to include an explicit re-identification and reconstruction risk assessment per dataset type and use case.

Quantitative or semi-quantitative re-identification risk assessments will be performed as part of the DPIA process, taking into account both data risk (content) and context risk (controls and environment). These assessments will rely on state-of-the-art guidance from Data Protection Authorities and relevant scientific literature on re-identification of medical imaging data.

2.3.4 Legal basis and secondary use

The primary collection and use of CT data at each Clinical Partner site are governed by the local legal framework and ethics approvals, and may rely on different legal bases under Article 6 and Article 9 GDPR (for example, provision of healthcare, public interest in the area of public health, or scientific research as implemented in national law).

For NetZeroAICT, the reuse of these data is considered a secondary use for scientific research purposes, compatible with the original purposes under the conditions described in the AICT DPIA and LIA and any site-specific ethics approvals. Depending on national law and the type of institution, Controllers may rely on a combination of:

- Article 6(1)(e) GDPR (task carried out in the public interest) or Article 6(1)(f) GDPR (legitimate interests), and

- Article 9(2)(j) GDPR (scientific research purposes) or Article 9(2)(i) GDPR (public interest in the area of public health), as implemented by national law.

The exact legal basis used by each Clinical Partner is documented at local level and reflected in the DPIA/LIA and in the AICT/NetZeroAICT legal framework. Pseudonymization is not used as a substitute for a valid legal basis; it is a complementary safeguard that reduces risk and facilitates multi-site research while maintaining compliance with GDPR.

2.3.5 Data subject rights in a Pseudonymized environment

Because only the Clinical Partner holds the mapping between Project Subject IDs and real-world identifiers, Data Subjects exercise their rights (access, rectification, erasure, restriction, objection and, where applicable, withdrawal of consent) primarily through the Clinical Partner that provided their data. The Clinical Partner, acting as Controller, is responsible for authenticating the request, assessing its validity and scope, and then, where appropriate, instructing CMRAD and the Joint Controllers to implement the requested measures in the Repository and downstream datasets.

The Platform supports this by:

- Storing data under stable Subject IDs that can be referenced in instructions from the Clinical Partner
- Enabling targeted deletion or restriction of cases linked to a given Subject ID, including in derived datasets where reasonably feasible and where this does not compromise scientific integrity beyond what is proportionate
- Maintaining audit logs of access and deletion events.

Because NetZeroAICT deliberately does not hold the re-identification key, it cannot itself directly identify individual patients and therefore cannot respond to Data Subject requests in isolation. This limitation is by design and is clearly explained in the DPIA, privacy notices and contractual framework. However, the combination of local re-identification capability at the Clinical Partner and technical tools in the Platform ensures that rights can still be effectively exercised and that instructions can be implemented across the Repository.

2.4 Will you re-use any existing data and how (including data protection aspects)?

Yes. The Project will primarily re-use existing CT datasets that have already been collected and stored in the Clinical Partners' PACS systems as part of routine clinical activity, rather than collecting new data solely for research purposes. This approach avoids unnecessary duplication of imaging examinations and is aligned with the principle that existing real-world data, when handled safely, can support high-impact research.
From a data protection perspective:

- The Clinical Partners, as Controllers, are responsible for ensuring that the primary collection and clinical use of the data comply with GDPR and national law, including appropriate information to patients and, where required, ethics approvals;
- The secondary use for NetZeroAICT is governed by the AICT DPIA and LIA and by site-level ethics approvals or local research governance decisions, which confirm that the reuse for scientific research is compatible with the initial context;

- Before any reuse occurs, data are Pseudonymized and minimized at the Clinical Partner site via CM-Connect, and only Pseudonymized Data are transferred to the Repository;
- The Project will not reintroduce direct identifiers, nor will it attempt to re-identify patients, in line with the contractual prohibitions and Platform terms of use.

The Project also envisages re-use of the curated Repository for future research projects where appropriate legal, ethical and contractual conditions are met. This will be governed, after the end of NetZeroAICT, by additional decisions of the Consortium and by updated governance instruments, ensuring continuity with the present framework.

The clinical images that constitute the fundamentals of this Project are acquired by each Clinical Partner for a primary purpose in order to e.g. diagnose, confirm a diagnosis or track disease or treatment progression. This data is stored within each hospital's picture archiving and communication system (PACS). For the performance of the Project, the data initially collected by each Clinical Partner, and stored in each PACS, will be transferred to the Project data Repository and **re-used in the Project, which will be processed for secondary purposes**.

In this regard the processing of data conducted on the Collective Minds Platform for the performance of the Project is considered a **secondary use of personal data**, so it is **conducted regardless of the initial legal basis** on which each Controller (e.g. institution, hospital, etc.) relies on for processing and collecting the data uploaded to the Platform.

In these cases, the use of pseudonymization techniques enables data processing for secondary purposes without the need to obtain the explicit consent of the Data Subjects.

Furthermore, whenever necessary, Collective Minds applies further anonymization techniques on the Pseudonymized datasets (in local environments) to obtain strongly Pseudonymized data that will enable safe further data processing minimizing the privacy risks of the Data Subjects to the maximum extent possible.

For the analysis of the risks involved with the sharing of data we are taking into account the ideas and recommendations provided in the Opinion 06/2013 on open data and public sector information (June 2013) (hereinafter 'PSI Opinion') reuse and the Opinion 05/2014 on Anonymization Techniques (April 2014) of the ARTICLE 29 DATA PROTECTION WORKING PARTY (hereinafter WP29), along with the Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU. This is because the recommendations for data sharing within the public sector encompass the highest standards to be taken into account for data sharing and processing for secondary purposes.

As reviewed in more detail in Opinion 3/2013 of the WP29 on purpose limitation, assessing whether further processing of personal data is incompatible with the purposes for which those data have been collected requires a multi-factor assessment. Account shall be taken in particular of:

(a) the relationship between the purposes for which the personal data have been collected and the purposes of further processing;

(b) the context in which the personal data have been collected and the reasonable expectations of the data subjects as to their further use;

(c) the nature of the personal data and the impact of the further processing on the data subjects;

(d) the safeguards applied by the Controller to ensure fair processing and to prevent any undue impact on the data subjects.

As the relationship between the purposes (a) and the main technical IT security safeguards (d) are described hereinabove. To safely and lawfully share and process the data, we identify two general sources of risk: **1)** The **context** (b) of the data sharing; and **2)** the **content** and/or **data** stored on the Platform, in which we need to analyze both the nature and the specific safeguards applied on the data to minimize the risk and the impact of the processing (c) and (d).

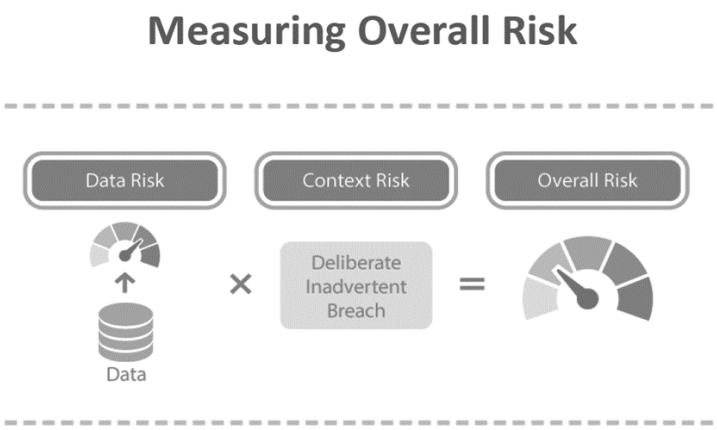

Figure 7: Diagram of the risk approach taking into account both the Data Risks and the Context Risks

The context represents the security, privacy, and contractual controls that are in place. For example, one context can be a public data release (e.g., an open data initiative) or, by contrast, a researcher who analyzes the data in a very secure enclave. These are two very different contexts and the risk of (re)identification is different in each of these, even for the same data.

As established by the WP29 in the PSI Opinion:

"[...] it is important to carefully consider what measures - including both legal and technical measures - could be put in place to help ensure that data protection concerns [...] will be addressed. It is particularly important to consider how re-users will access the data [...]. In this respect it is crucial what additional security controls will be put in place, such as, for example, a [...] verification system to prevent automated access and minimise the risk of harvesting an entire database. Using specific technical measures could help reduce misuse of personal data and negative impacts on data subjects that otherwise could be made possible by unlimited and unconditional access of reusers to entire datasets [...]".

Furthermore, in the PSI Opinion the WP29 recommends that:

The terms of the license to re-use [...] should include a data protection clause, whenever personal data are processed including also situations where anonymized data sets derived from personal data will be made available for reuse;

Where appropriate, public-sector bodies should ensure that personal data are anonymized, and license conditions specifically prohibit re-identification of individuals and re-use of personal data for purposes that may affect the Data Subjects;

As described above, Collective Minds applies legal and technical measures to the following context:

- Facilitating and enhancing collaboration between health professionals, professors (and students), physicians (medical doctors), and researchers is necessary and desirable. Furthermore, this collaboration must include the sharing of knowledge of Real-World Data that enhances the knowledge and provides real impact on those professionals' activity knowledge and training.

The main contextual risks are:

- The sharing of sensitive data, such as health information.
- Having no control over who has access to the data and what further processing activities are carried out on it (processing for secondary purposes that are unrelated to the primary purposes).
- An attempt to identify or re-identify a Data Subject.
- The processing of personal data for malicious, unlawful, and/or unethical purposes, such as identifying the patient or uncontrolled data processing.

As a result, the Consortium, through Collective Minds, takes the following measures:

- A closed Platform, where only registered users that are part of the research Consortium can access and process the datasets: The Collective Minds Platform.
- Establishing technical and operational measures on the Collective Minds Platform to avoid unauthorized access, preventing rogue/malicious behaviour or data processing outside the controlled environment of the Platform.
- Control who can register as a User of the Collective Minds Platform through a vetting procedure to verify that only trusted and relevant healthcare professionals, professors, physicians (medical doctors), and researchers can be registered as users of the Platform.
- Full traceability and access control.
- Users must sign and accept legal undertakings (Terms and Conditions) before registering on the Platform. The Terms and Conditions include, among other things:
    - A confidentiality / non-disclosure agreement regarding the data and the content of the Platform, including personal data and or anonymized data and any intellectual property rights.
    - There are several data notices and a privacy policy that must be read, followed, and accepted.
    - To ensure that the user fully understands its obligations beyond the terms and conditions and general terms, information to the user is provided, generally in the form of warnings, informative messages, and guidelines.
    - The prohibition of sharing or processing data outside the Platform control and traceability.
    - The prohibition of processing data for unauthorized, unlawful or unethical purposes.
    - It is prohibited to attempt to or identify a Data Subject, or to conduct any activities of re-identification.
    - It is possible to legally enforce any of the above measures if they are violated.

Aside from the contextual risks described above, technical and organizational measures are implemented on the **Platform's data (content)**, before being transmitted and stored in the Repository, in order to avoid a potential or actual impact of the rights, freedoms, and dignity of Data Subjects.

As provided by the WP29, in the Opinion 05/2014 on Anonymization Techniques (April 2014)*: "'Open data' may provide clear benefits for society, individuals and organizations, but only if everybody's rights are respected to the protection of their personal data and private life".*

Before entering into the analysis of the measures regarding the data risks (content), it is important to clarify that a health dataset will have two types of variables that we care about from a privacy perspective: (a) direct identifiers, and (b) indirect identifiers.

Direct identifiers are details such as names, addresses, social insurance numbers, telephone numbers, email addresses, and any other unique identifiers. These direct identifiers are typically removed from the dataset when it is shared for secondary purposes. The unique identifiers, such as a medical record number, would be converted to a pseudonym so that it can still be used to relate all of the records that belong to the same patient. When you do all of these things to protect against re-identifying individuals from these types of variables, the dataset is considered Pseudonymized.

As a result of the procedure described above, the Collective Minds Platform does not collect, process, store, or share any direct identifiers, as it is designed as a technical system that ensures all data uploaded to the Platform is strongly Pseudonymized before transmission.

Therefore, **this make secure processing for secondary purposes possible** since, following the latest recommendations from the EU Data Processing Authorities, the data is processed through a 'proxy' function (CM Connect) that enables to perform and apply operations and measures on the data, within the safe and controlled environment of each data Controller before uploading the data into the Collective Minds Platform.

2.5. What is the origin of the data?

The Repository will receive CT data from the AICT Consortium's Clinical Partners, which include institutions in Europe and beyond. Each hospital group contributes an existing CT archive in a pathology-agnostic fashion, covering multiple anatomical regions and clinical indications.

The imaging data are therefore:

- Originally collected as part of routine care in hospitals or imaging centres
- Stored in local PACS infrastructures under the responsibility of each Clinical Partner as Controller
- Then Pseudonymized and transferred via CM-Connect to the NetZeroAICT Repository under the Controller–Processor relationship described above.

The resulting Repository captures a diverse set of populations, disease contexts and imaging protocols, making it a valuable resource both for NetZeroAICT and for potential future research under appropriate governance.

The NetZeroAICT Repository will receive CT data from and international Consortium (AICT Consortium, which includes clinical sites from: France, Greece, Poland, Belgium, England, Scotland, Australia and Brazil), see Table 1 for further information. Each of these hospital groups will share their existing CT clinical archive in a 'pathology agnostic' manner. The Repository will therefore capture the most diverse geographic, populational and disease context as compared to the other research imaging datasets in existence. We will have 1 million cases in the NetZeroAICT Repository. The imaging data will be provided by the Clinical Partners of the Project listed in Table 1.

Table 1. List of Clinical Partners providing data to the NetZeroAICT Consortium

Clinical partner	Country
GZA	Belgium
Poznan	Poland
Heraklion	Greece
USP	Brazil
Glasgow	UK
Nice	France
AZ Sint-Jan	Belgium
UFF - Universidade Federal Fluminense	Brazil
Leicester	UK
Healius	Australia

2.5. What is the expected size of the data?

The Project aims to include a cohort of up to 500,000 patients in the initial stages, with milestones as follows:

- D2.3: 100,000 datasets transferred to the Repository (M6)
- D2.4: 500,000 datasets transferred to the Repository (M24)
- D2.5: 1,000,000 datasets transferred to the Repository (M36).

For the purposes of storage estimation, each CT examination (for example, one CE series and one non-CE series) is treated as a dataset. The average dataset size is estimated at around 0.5 GB. This implies a base data volume in the range of:

1,000,000 datasets × 0.5 GB ≈ 500 TB by month 36.

Additional storage will be required for derived datasets, AI model outputs and backup copies, but these can be managed using scalable cloud infrastructure.

2.6. To whom might it be useful ('data utility')?

The data collected and curated in NetZeroAICT will be useful for:

1. The NetZeroAICT Consortium Partners, who will use the Repository to generate scientific results, develop and validate AI models, and create new knowledge on environmentally sustainable imaging
2. The broader AICT community, given that NetZeroAICT builds on and extends the AICT data infrastructure and governance model
3. Future research initiatives, which may leverage the Repository under appropriate governance to develop or evaluate other AI tools in CT imaging
4. Ultimately, patients and society, by enabling reduced RCM usage, lower environmental impact and improved patient experience (for example, avoiding contrast injections where clinically safe), while maintaining or improving diagnostic quality.

3. FAIR data

3.1. Overview

This Project will generate and process medical imaging data in compliance with the FAIR principles, ensuring that all data are Findable, Accessible, Interoperable, and Reusable throughout the Project lifecycle and beyond. The data will primarily consist of DICOM-formatted medical images and its associated metadata obtained from partner healthcare institutions.
All activities will comply with EU data protection regulations (GDPR), institutional ethics approvals, and relevant standards for medical data management and anonymization.

3.2 Data description

The dataset will consist of imaging studies (CT) acquired in DICOM format, following the standard defined by ISO 12052:2017. Each DICOM object contains both pixel data and a structured metadata header, which includes acquisition parameters, patient demographics (anonymized), and study-level identifiers.

Derived datasets may include:

- Processed or AI-enhanced DICOM series (Digital contrast)

- Quantitative image biomarkers or structured reports (Radiological reports). This is to be discussed within the Consortium.
- Metadata extracted from DICOM headers and stored in tabular or JSON formats.

All datasets will be stored with consistent, documented identifiers and metadata schemas.

3.3 Ensuring FAIR Data Principles

3.3.1 Overview

NetZeroAICT manages data according to the FAIR principles, while recognizing that the core dataset consists of Pseudonymized CT imaging data that remains personal data under the GDPR and therefore cannot be made fully open. The Project follows the principle “as FAIR as possible, as closed as necessary”.

Data will be stored and processed in the secure Platform operated by CMRAD, primarily in DICOM format with associated structured metadata. Within this controlled environment, data will be made findable and accessible to authorized users from the Consortium, interoperable through the use of standards and controlled vocabularies, and reusable through clear documentation and governance.

All FAIR-related activities will be carried out in compliance with the GDPR and other applicable data protection laws, as well as institutional ethics approvals and relevant standards for medical data management and pseudonymization. Where additional aggregation or transformation allows for truly anonymized or negligible-risk outputs to be produced, these may be shared more broadly under appropriate licenses; this will be decided on a case-by-case basis and documented in future updates of the DMP.

Data description

The primary dataset will consist of CT imaging studies acquired in DICOM format, following ISO 12052:2017. Each DICOM object contains image pixel data and a structured header. Before transfer to the Repository, direct identifiers are removed and quasi-identifiers are minimized or generalized at the Clinical Partner site through CM-Connect, so that the DICOM metadata in the Repository contain only Pseudonymized and minimized information such as modality, body region, scanner parameters, and limited, pre-defined demographic or clinical descriptors where strictly necessary.

In addition to the raw CT series, the Project will generate derived datasets. These may include Digital Contrast CT series obtained through AI processing; structured, machine-readable summaries of image content or radiological findings where the Consortium decides to include such outputs; and metadata extracts from DICOM headers and curation processes stored in tabular or JSON formats inside the Platform’s databases. All datasets in the Repository will be associated with consistent internal identifiers (for example Site ID, Subject ID, Study ID, Series ID) and will follow documented metadata schemas defined in WP2 and WP3.

3.3.2 Findable

Within the Repository, imaging data and associated metadata will be indexed in a dedicated cohort and metadata database. Each case will be identified by a stable Subject ID and by exam- and series-level identifiers. Metadata will follow a standardized structure based on DICOM fields, enriched with Project-

specific descriptors such as anatomical region, contrast status (CE / non-CE), scanner vendor and model, and high-level protocol characteristics relevant for AI development.

Provenance information (for example, contributing Clinical Partner and Pseudonymization pipeline used) will be recorded in the metadata. Detailed operational timestamps and free-text identifiers that increase re-identification risk will not be stored in the Repository; instead, only the information needed for scientific utility and traceability at Project level will be retained.

Inside the Platform, researchers will be able to search and filter cases using these metadata (for example, by body region, contrast status, scanner type or site) in order to define cohorts. For any datasets or derived resources that are deposited in external repositories, persistent identifiers such as DOIs may be assigned, together with a description of the dataset and its limitations, to improve long-term findability outside the Platform.

3.3.3 Accessible

Accessibility is organized around a closed, secure Repository rather than unrestricted public access. Authorized users from the Consortium will access the data via authenticated interfaces (web application and API) with role-based permissions. Access rights will be configured in accordance with the user's role, institutional affiliation and Work Package responsibilities, and all access will be logged.

Imaging data will be stored in DICOM with lossless compression to ensure that once access is granted, the data can be retrieved efficiently and without loss of diagnostic information. Only Pseudonymized Data will be accessible in the Repository; direct identifiers and the mapping between Subject IDs and real-world identities remain solely at the Clinical Partner site. There will be no direct access within the Platform to identifiable patient data.

At this stage, core CT datasets are not intended to be made openly accessible outside the Consortium because of the residual re-identification and reconstruction risk. However, high-level aggregate information about the Repository (for example, counts of cases per body region, site or protocol category) may be published to inform the scientific community and to increase transparency. In later phases of the Project, and subject to updated DPIA results and Consortium agreement, the possibility of providing controlled external access to selected subsets or derived data will be re-evaluated.

3.3.4 Interoperable

Interoperability is achieved by combining standard technical formats with harmonized semantics. DICOM is used as the standard format for image data; associated metadata and derived variables are stored in database tables that can be exported, when appropriate, as CSV or JSON for use in analysis environments or other systems.

Where feasible, the Project will map metadata and clinical descriptors to controlled vocabularies and ontologies. Examples include RadLex for imaging concepts and procedures, and SNOMED CT or ICD families for diagnoses and indication categories, taking into account licensing and practical constraints. WP2 and WP3 will maintain a Project-wide data dictionary that describes variable names, value sets and mapping to these standards, ensuring consistent interpretation across sites and Work Packages.

The Platform will expose internal APIs to allow authorized tools and processes to query metadata, define cohorts and access Pseudonymized Data inside the secure environment. During the Project, the

Consortium will also explore, subject to time and resources, the feasibility of exposing selected metadata through standardized interfaces aligned with HL7 FHIR Imaging Study or similar profiles, in order to facilitate integration with other infrastructures and future reuse.

3.3.5 Reusable

Reusability depends on clear documentation, stable identifiers, and a well-defined legal and governance framework. For each dataset and cohort in the Repository, the Platform will record key information on data origin, inclusion criteria, acquisition parameters, Pseudonymization and preprocessing steps, and curation status. This information will allow other Work Packages and, in the longer term, future projects operating under the same governance, to understand how the data were generated and to assess whether they are suitable for specific analytical purposes.

The Repository will incorporate a labelling system that allows datasets or sub-cohorts to be earmarked for particular functions (for example, dedicated training sets, held-out validation or test sets, or cohorts reserved for external benchmarking). This labelling supports rigorous experimental design and prevents unintended reuse of the same cases across training and validation phases.

Any datasets or derived outputs that are released outside the Consortium will be accompanied by explicit license and usage conditions. For genuinely anonymized or negligible-risk aggregated outputs, standard licenses such as Creative Commons (for example, CC-BY or CC-BY-NC) may be used. For software components or trained models, appropriate open-source licenses may be selected where compatible with the Partners' exploitation plans. All such decisions will be documented in future updates of the DMP, together with an explanation of how re-identification and reconstruction risks were assessed and mitigated for the shared resource.

Long-term preservation of Pseudonymized Data and associated metadata will be ensured through storage in the CMRAD infrastructure, with secure backups and redundancy, for at least ten years after the end of the Project or for the period required by applicable legal and contractual obligations.

4. Allocation of Resources

4.1 Costs for making data FAIR

Making data FAIR in NetZeroAICT requires:

- Development, operation and maintenance of the Platform and Repository
- Implementation and monitoring of Pseudonymization and data minimization pipelines at Clinical Partner sites
- Curation, harmonization and quality control of metadata and any associated clinical variables
- Legal, ethical and governance work to maintain the DPAs, JCA, DPIA, LIA and related frameworks.

These activities are integrated into the budgets of WP2 and related Work Packages and are considered an essential part of the scientific and technical work, rather than an optional add-on.

4.2 Coverage of costs

Costs related to data management and FAIR implementation are covered through:

- The Horizon Europe funding allocated to WP2 and to other WPs with data-intensive tasks (for example, WP3, WP4, WP5)
- In-kind contributions and effort from Clinical Partners, who designate local staff to oversee Pseudonymization, data selection and local governance
- CMRAD's infrastructure and operational costs for the Platform, including cloud hosting, security, backup and monitoring.

Where additional costs arise for long-term preservation or external repository deposition of aggregated outputs, these will be planned within the Project's budget and, if needed, included in future amendments or sustainability plans.

4.3 Responsibility for data management

Overall responsibility for data management lies with the WP2 lead (CMRAD), acting in close collaboration with:

- The Project Coordinator
- The leaders of data-intensive Work Packages (WP3, WP4, WP5)
- The legal and ethics leads who oversee data protection, DPIA and ethics compliance.

Each Clinical Partner appoints a local contact or data steward responsible for:

- Overseeing CM-Connect deployment and operation
- Ensuring that local Pseudonymization and minimization follow agreed procedures
- Coordinating with local data protection officers and ethics bodies.

5. Data security, personal data breach, and incident management

5.1. What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

The Platform operated by CMRAD follows industry-standard security practices and is designed as a secure processing environment. Key controls include:

- Storage of imaging data and metadata in secure cloud services within the EU (such as AWS S3 and AWS Glacier), with redundancy and regular backups
- Encryption of data in transit (for example, TLS for all connections between CM-Connect and the Platform) and at rest
- Strict access control mechanisms, including role-based permissions, multi-factor authentication, and segregation of duties
- comprehensive logging of user actions and administrative operations, enabling traceability and audit
- Regular security updates, vulnerability management and monitoring of the underlying infrastructure.

These measures are complemented by the design choice that direct identifiers never leave the Clinical Partner environment, and that only Pseudonymized Data are present in the Repository.
Disaster recovery and business continuity procedures are in place to ensure that data can be restored in case of infrastructure failures, while preserving integrity and confidentiality.

5.2. Is the data safely stored in certified repositories for long term preservation and curation?

Yes, data is safely stored in AWS S3 and AWS S3 Glacier that follows up to date standards such as SOC 1, SOC 2, SOC 3, ISO 27001, ISO 27017, ISO 27018, and PCI DSS.

The Repository is hosted in cloud services that comply with relevant security and certification standards (for example, ISO 27001, ISO 27017, ISO 27018 and SOC reports), ensuring that data are stored in an environment with appropriate technical and organizational safeguards.

Long-term preservation of Pseudonymized Data within the Platform will be supported for at least the duration of the Project and, subject to Consortium decisions and legal constraints, may be extended beyond the Project's end to enable continued research under appropriate governance.

5.3 Personal Data Breach response and incident management

The NetZeroAICT Platform and Repository are designed to minimise the likelihood and impact of any Personal Data Breach, but the Consortium recognises that incidents cannot be ruled out. A structured incident-management process is therefore in place, aligned with the GDPR (in particular Articles 33 and 34), the Joint Controller Agreement (JCA), and the Data Processing Agreements (DPAs) between Clinical Partners and CMRAD.

A **Personal Data Breach** is understood as any breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, Personal Data transmitted, stored or otherwise processed in the Project.

5.3.1 Detection, internal reporting, and initial containment

CMRAD continuously monitors the Platform and Repository for security-relevant events, including unusual access patterns, failed authentication attempts, anomalous behaviour of services and integrity checks on stored data. When an event suggests that a Personal Data Breach may have occurred, CMRAD's internal incident response procedure is triggered.

The first objective is to contain and limit the impact of the suspected breach. Depending on the nature of the event, this may involve isolating affected systems or services, temporarily suspending specific user accounts or access paths, and applying emergency configuration changes. All such measures are documented, and the relevant Technical and Security teams at CMRAD are involved from the outset.
Users of the Platform are instructed, through internal documentation and contractual terms, to report any suspicious behaviour, inadvertent disclosure, or suspected compromise immediately to CMRAD and to their own institutional contact points.

5.3.2 Assessment, classification, and notification

Once initial containment steps are in place, CMRAD performs a structured assessment to determine whether the event constitutes a Personal Data Breach affecting data in the NetZeroAICT Repository, and, if so, its likely scope and severity. This assessment will consider, in particular:

- Which components of the Platform were affected
- Which categories of Personal Data and Pseudonymised Data were involved
- The number and nature of Data Subjects potentially impacted
- Whether data were merely made temporarily unavailable, or whether integrity and confidentiality were compromised
- Whether the breach is likely to result in a risk or high risk to the rights and freedoms of Data Subjects.

As Processor, CMRAD has a contractual obligation to notify the relevant Controllers and Joint Controllers without undue delay once it becomes aware of a Personal Data Breach affecting the Repository. Such notifications are sent to the designated contacts of the Clinical Partners and, where appropriate, to the governance bodies of the Consortium. They include, to the extent possible at that moment, a description of the nature of the Breach, the categories and approximate number of Data Subjects and data records concerned, likely consequences, and proposed or implemented measures to address the Breach.

Each Clinical Partner, acting as Controller, remains responsible for assessing whether the Breach triggers notification obligations to its competent Supervisory Authority and, where applicable, to Data Subjects, in accordance with Articles 33 and 34 GDPR or corresponding local provisions. CMRAD supports Clinical Partners by providing the information and logs reasonably required for those assessments.

If an incident affects only local systems at a Clinical Partner site, before pseudonymisation and transfer via CM-Connect, the Clinical Partner leads the incident response as Controller of those data. In such cases CMRAD will cooperate if there is any indication that credentials, interfaces or trust relationships involving the Platform may also have been impacted.

5.3.3 Communication, remediation and lessons learned

Where notification to a Supervisory Authority or to Data Subjects is required, the content and channels of such communication are determined by the Controllers and Joint Controllers, considering legal requirements, the nature of the Breach and the safeguards already in place (for example, pseudonymisation and access controls). CMRAD may be asked to provide technical input or to contribute to explanations of the Platform's role and safeguards.
In parallel, CMRAD performs a detailed root cause analysis to identify how and why the Breach occurred and which technical or organisational controls failed or proved insufficient. Corrective and preventive measures are then defined and implemented, such as configuration changes, additional access restrictions, software fixes, process changes, further training or, where necessary, architectural adjustments.
The outcome of significant incidents, together with the remediation steps, is reported to the relevant Consortium governance bodies. Where the Breach reveals a change in the risk landscape (for example, new plausible attack vectors, novel reconstruction risks, or weaknesses in contextual controls), the DPIA and related risk assessments are reviewed and updated. Any resulting changes in security measures, data handling procedures or access policies are reflected in future versions of the DMP and in the Platform documentation.
By combining clear detection and reporting channels, a structured assessment and notification workflow, and a commitment to remediation and learning, NetZeroAICT aims to ensure that any Personal Data Breach is handled promptly, transparently and in full alignment with the responsibilities of Controllers, Joint Controllers and Processor under the GDPR and related agreements.

6. Ethical aspects

There is a dedicated team that represents 1) legal, privacy and data protection 2) Ethics and 3) patient engagement, that is devoted to specifically analyse, oversee and ensure compliance with the ethical standards, assess ethical risks and minimize them.
This taskforce will perform ethics and fundamental rights impact assessment during the performance of the Project. Furthermore, an Ethics Committee will also be in charge of ensuring ethical aspects are duly considered.
The Project processes Special Categories of Personal Data on a large scale and therefore has an inherent ethical dimension. Ethical oversight is ensured by:

- Local ethics approvals or notifications obtained by Clinical Partners for the secondary use of CT data in NetZeroAICT, where required under national law
- A dedicated task force within the Consortium dealing with legal, data protection and ethics matters
- An Ethics Committee or equivalent body that reviews and monitors ethical issues throughout the Project.

The DPIA and LIA were performed in AICT and extended to NetZeroAICT, providing a structured assessment of risks to fundamental rights and freedoms and the mitigating measures applied. Any significant changes in data processing, new categories of data or new types of secondary use will trigger a review of these assessments and, where needed, updated mitigations.
The ethical framework explicitly addresses:

- The balance between environmental benefits, scientific value and risks to individuals
- The use of Pseudonymization and a closed processing environment rather than open data release
- The mechanisms for respecting Data Subject rights in a Pseudonymized setting
- Transparency towards patients and the public, through appropriate information materials produced by Clinical Partners and the Consortium.

7. Conclusions

This first version of the NetZeroAICT Data Management Plan identifies the datasets that will be collected, re-used or generated by the Project, describes the technical and organizational framework for their management, and explains how data protection, ethics and FAIR principles are integrated into the design.

Key elements include:

- A clear allocation of roles and responsibilities among Controllers, Joint Controllers and Processor
- A strong Pseudonymization and minimization strategy implemented at Clinical Partner sites via CM-Connect
- Explicit recognition of residual re-identification and reconstruction risk and a commitment to state-of-the-art risk assessment
- A closed, secure processing environment that supports large-scale AI development while limiting data leakage risk
- A realistic and risk-based approach to openness, focusing on sharing of aggregated or derived outputs rather than raw CT data.

This version will be updated as part of D2.2 (month 24), taking into account the practical experience from data ingestion and curation, the outcomes of more detailed risk assessments and any new legal or ethical requirements.

8. Degree of progress

The deliverable is fully achieved for this first version. Data management procedures are in place to support the initial ingestion and curation milestones, and the governance framework is aligned with the legal and ethical instruments established in AICT and updated for NetZeroAICT. An Updated Data management plan will be submitted as D2.2 by M24.

9. Dissemination level

This document is foreseen as a public deliverable, subject to the constraints imposed by the Grant Agreement and the Consortium Agreement. Any future changes to its dissemination level will be decided by the Consortium in line with those instruments.