late aya

LATE-AYA

Understanding and addressing LATE-effects of treatment of AYA cancer survivors with AI-based digital phenotyping and non-invasive holistic approach

Project No.101214326

Deliverable

D1.4 Data Management Plan

29/11/2025

Co-funded by the European Union

Co-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 European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

Document Details	
Project Name	Understanding and addressing LATE-effects of treatment of AYA cancer survivors with AI-based digital phenotyping and non-invasive holistic approach
Project Acronym	LATE-AYA
Grant Agreement Number	101214326
Start Date of the Project and Duration	1 June 2025; 60 months
Project Coordinator Name	Universidad Politécnica de Madrid (UPM)
Task	T1.4
Deliverable No.	D1.4
Title of Deliverable	Data Management Plan
Beneficiary Responsible	UDG Alliance
Authors	Adrian Quesada Rodriguez, Renáta Radócz, Natálie Binderová, Sébastien Ziegler, Ana María Pacheco (UDGA)
Reviewers	UDEU, INT
Status (draft, final)	Final
Dissemination level (PU, SEN)	PU
Keywords	Data Management Plan, FAIR principles, Regulatory compliance, IPR management, Data Standards, Interoperability

Version history		
Version	Date	Description
V0.4	01/08/2025	Initial draft by UDGA
V0.8	20/11/2025	Input request (questionnaires) from all partners.
V0.9	28/11/2025	Peer review by UDEU & INT completed
V1.0	29/11/2025	Final verification by UPM

Table of Contents

This project has received funding from the European Union's EU4Health Programme under Grant Agreement No 101214326

List of tables

This project has received funding from the European Union's EU4Health Programme under Grant Agreement No 101214326

List of abbreviations

AYA	Adolescents and Young Adults
AI	Artificial Intelligence
AI Act	Artificial Intelligence Act
AYA	Adolescents and Young Adults
CC 0	Creative Commons Zero
CC BY	Creative Commons Attribution license
Dno.no	Deliverable
DMP	Data Management Plan
DSA	Digital Services Act
EHDS Regulation	European Health Data Space Regulation
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FHIR	Fast Healthcare Interoperability Resources
GA	Grant Agreement
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
LE	Late Effects
MDR	Medical Device Regulation
ML	Machine Learning
NIS2	NIS 2 Directive
OMOP	Observational Medical Outcomes Partnership
PREM	Patient-Reported Outcome Measure
PROM	Patient-Reported Outcomes Measures
SPE	Secure Processing Environment
ZTA	Zero Trust Architecture

Abstract

This first version of the Data Management Plan (D1.4) presents a data summary, including types, sizes, origins, purposes, and formats of data that are collected, generated, and reused within LATE-AYA. Subsequently, it outlines how FAIR principles (Findable, Accessible, Interoperable, Reusable) apply to the project's data, metadata, software, and other research outputs. Additionally, it provides an overview of data security measures, including information on data recovery and secure repositories.

D1.4 is preparatory and precedes several core legal and architectural documents required before clinical activities begin. To this end, it establishes the legal and technical requirements necessary for subsequent foundational deliverables, including the Framework Architecture (D2.4), the Detailed User Needs Specification (D2.1), and crucially, the Data Agreements Templates (D2.3) (due M24).

The plan also maps out regulatory considerations that partners may encounter during the project's duration. It examines obligations under the General Data Protection Regulation (GDPR), the European Health Data Space (EHDS) Regulation, the AI Act, the NIS2 Directive, the Cyber Resilience Act, and the Digital Services Act. This overview is intended to serve as a starting point for compliance. Moreover, the plan highlights key ethical issues, such as the principles governing Intellectual Property Rights (IPR) and Access Rights to Foreground results and Background data, which are specified in the CA, to maximize scientific and economic return while fulfilling open science mandates.

Particularly, this document touches upon the principles associated with the inclusion of minors and the management of ethical consent, which are of relevance as minors are likely to be included in the target group for LATE-AYA's clinical studies. In this sense, it seeks to establish a baseline for the specific ethical safeguards that must be implemented, including strict informed consent and assent processes, ongoing supervision by ethics experts, and adherence to European international guidelines, such as the European Convention on Human Rights, the Declaration of Helsinki, and the UN Convention on the Rights of the Child. This deliverable note this as a reference to the corresponding tasks in other WPs (Particularly WP2) to ensure alignment between planning, management, compliance and implementation activities.

This DMP lays the groundwork for secure, standardized, and ethically responsible data management and for the future deliverables in LATE-AYA. It supports the project's contributions to European research networks like UNCAN.eu and directly aids the EU Mission on Cancer and Europe's Beating Cancer Plan by promoting data-driven and equitable survivorship care across Europe. It details the use of relevant health data interoperability standards such as OMOP (Observational Medical Outcomes Partnership) and FHIR (Fast Healthcare Interoperability Resources)

To this end, the plan also introduces the contents of the common DMP chapter generated by the "QoL AYA" cluster as an annex, and will continue to align its upcoming activities with associated developments.

About this document

Introduction to the Late-Aya Project

The LATE-AYA (Understanding and managing LATE-effects of treatment of AYA cancer survivors by AI-based digital phenotyping and non-invasive holistic approach) project focuses on addressing the distinct health and psychosocial challenges faced by adolescent and young adult (AYA) cancer survivors, typically diagnosed between the ages of 15 and 39. Even though onco-developments have enhanced survival rates, most survivors suffer from significant late effects (LE) from treatment, including secondary cancers, cardiovascular diseases, sterility, and mental disorders like anxiety and depression. LATE-AYA aims to address a significant gap in survivorship care by empowering AYAs to optimise their health and well-being in the post-cancer period. The objective of this initiative is to develop a privacy-preserving, interoperable, and patient-centred digital ecosystem for personalised follow-up, with a focus on preventive health behaviours, psychological support, and social reintegration.

This objective is realised through the integration of data from multiple sources, including conventional clinical records, continuous data from personal digital devices such as smartphones and wearables, and Patient-Reported Outcome and Experience Measures (PROMs and PREMs). The management of data will be conducted within a cloud-based Secure Processing Environment (SPE), which has been constructed on the principles of Zero Trust Architecture (ZTA). This ensures that the personal data is subject to maximum protection. The secure infrastructure is fully compatible by design with existing General Data Protection Regulation (GDPR) principles and the forthcoming European Health Data Space (EHDS) Regulation specifications, ensuring compliance, data sovereignty, and ethical control throughout the project.

In line with these aims, LATE-AYA focuses on the following key objectives:

- To collaboratively design and develop a safe, privacy-sensitive digital space for the integration and analysis of AYA cancer survivor clinical, behavioural, and psychosocial information, involving patients, caregivers, and clinicians actively in its development.
- To develop a cloud-based Secure Processing Environment (SPE), founded on Zero Trust Architecture (ZTA) principles, to facilitate ethical data collection and support advanced, AI-based analysis of clinical data for research.
- To enable data managed according to FAIR principles (Findable, Accessible, Interoperable, and Reusable) by applying defined standards like OMOP (Observational Medical Outcomes Partnership) and FHIR (Fast Healthcare Interoperability Resources), therefore facilitating clean reuse and integration of datasets and models with European research infrastructures, in particular UNCAN.eu.

This project has received funding from the European Union's EU4Health Programme under Grant Agreement No 101214326

- In support of individualized survivorship care through the use of AI-based digital phenotyping to model health trajectories, predict LE risk, and facilitate survivors and their care teams with early detection resources and individualized, whole-person interventions.
- To conduct a multicentre, two-stage trial: Stage 1 will be an observational study with at least 100 AYA survivors of cancer to establish a benchmark, and Stage 2, a randomised trial with 300–350 survivors to follow up for at least 18 months to assess the effectiveness of the digital intervention.
- To put in place effective knowledge-sharing and exploitation channels in order to disseminate findings to survivors, clinicians, researchers, and policymakers, and develop business plans for exploitable components and proactively cluster with similar EU initiatives, e.g., "Quality of life (AYA)" cluster.
- To include a firm ethical and legal basis that protects individual data and guarantees all activities in the project comply with the highest ethical requirements and European legislation, such as GDPR and the new EHDS Regulation.

Through a combination of clinical and digital expertise and survivor participation, LATE-AYA will leverage the potential of cost-effective, evidence-based cancer survivorship solutions via early Health Technology Assessment. The project aims to provide the European research infrastructures, such as UNCAN, with first-class FAIR datasets and certified AI models. This directly benefits the key European policies, such as the EU Mission on Cancer and Europe's Beating Cancer Plan, as well as supports advancing data-driven care for young cancer survivors in Europe in alignment with the long-term visions of these initiatives.

Purpose of this Document

As per the project objectives, this document aims to ensure that the personal data collection, storage, and usage in the LATE-AYA project are thoroughly documented. LATE-AYA will collect vast amounts of data, including sensitive personal information, through smartphones and wearable devices (smartwatches or rings). Consequently, several instances of this document are envisaged as the project is currently at that stage. For example, the “Data management plan v1” (D1.4) is due at month 6, and the “Data management plan v2” (D1.5) is due at month 58.

As an EU-funded scientific research project, all consortium members of LATE-AYA project are subject to the laws and regulations on data protection (e.g., the General Data Protection Regulation or GDPR for short). Regulatory compliance has to be documented to avoid legal consequences and ensure the safety of individual project participants, who are adolescents and young adults (AYA) as a vulnerable population. The project will adhere to the highest ethical standards by ensuring that all activities conform to both national and international legislation.

The DMP forms part of the quality and risk management strategy (task T1.4), balancing the project's compliance efforts with the need to respect open data and open research requirements, including FAIR principles as per Directive 1024/2019 (Open data Directive). The plan also envisages a common chapter for the "Quality of life (AYA)" cluster to harmonise data standards and data protection requirements, enabling data sharing with UNCAN.eu.

Overall, the aim is to ensure that LATE-AYA consortium members handle both sensitive and non-sensitive data appropriately and adequately document compliance measures. Lastly, it must be noted that the information provided in this document is based on the answers from the Data Management Questionnaires provided by the project beneficiaries, together with the information from the Grant Agreement.

Context of this Deliverable

As already mentioned, the present document is the first version of DMP entitled "Data management plan v1" (D1.4) due in month 6. It will reflect upon the activities initially performed by partners based on the project's stage in the first six months of its operational activity.

The information in this first version is meant to lay down the very basic principles of sharing data within LATE-AYA and between Consortium partners. This DMP is a part of the project's quality and risk management strategy (Task T1.4). It will also set the basis for the other key foundational outputs, such as "Detailed user needs specification" (D2.1), "Data agreements templates" (D2.3), and "Framework architecture" (D2.4). These deliverables will provide guidance and establish procedures for use by the project partners.

Future versions of the DMP will investigate data-related activities in more detail and provide additional information concerning specific datasets collected and/or generated by LATE-AYA. As a result, the DMP will be constantly updated during the project runtime. Nevertheless, the second deliverable "Data management plan v2" (D1.5) is due for delivery in month 58.

This project has received funding from the European Union's EU4Health Programme under Grant Agreement No 101214326

1. Data Summary

Data lies at the heart of the LATE-AYA project, serving as the core component for developing and implementing its AI-driven digital phenotyping platform. The main aim of the project is to understand and manage the late effects (LE) of cancer treatment in Adolescents and Young Adults (AYAs), empowering them to improve their health, well-being, and long-term quality of life.

As a result, the project relies heavily on collecting considerable amounts of quality data from AYA cancer survivors. This encompasses both "conventional data" types (e.g., clinical, psychosocial instruments, and Patient-Reported Outcome Measures) and "unconventional data" obtained non-invasively via personal digital devices such as smartphones and wearable devices. This enables the quantification of an individual's phenotype in real time and in natural environments on a continuous moment-by-moment basis. The following sections will delve deeper into the exact datasets that will be collected, the purpose they will serve, their source and method of collection, as well as their format.

a. Types of Data within Late-Aya and its Origin

As outlined above, LATE-AYA will implement an AI-driven digital phenotyping platform and supporting digital tools to monitor late effects of cancer treatment in approximately 400–450 AYA cancer survivors, aged 15–39, treated in hospitals across three EU countries. The study will be conducted in two stages: Stage 1, an observational study with at least 100 AYA cancer survivors, and Stage 2, a randomized trial with 300–350 survivors who will be followed for a period of 18 to 30 months. The LATE-AYA project is centered on re-using already existing data and collecting and/or generating new stakeholders', non-personal and personal data, the latter of which is considered highly sensitive due to its health-related nature and the vulnerability of the participants.

PERSONAL DATA
The project will collect 2 main types of personal data:

1. Conventional Data: this data will be collected during clinical follow-up visits and through questionnaires from all study participants.

    a. Demographic and Clinical Information:
        - Clinical information such as age, sex/gender
        - Socioeconomic characteristics/condition
        - Cancer history including:
        - Histology, staging, and tumor site
        - History of anti-cancer treatments (surgery, radiation, systemic medications), including dosage and timing
        - Medication and medical events review
        - Baseline clinical evaluation of physical/emotional/social status

through SoC visits

b. Patient-Reported Outcome Measures (PROMs) and Questionnaires[1]:
    - Health-Related Quality of Life questionnaires, including EORTC and SURV100
    - Other PROMIS measures that may be considered, including:
        - Distress Thermometer
        - Illness Perception Questionnaire-Revised
        - Fertility Problem Inventory
        - Reproductive Concerns After Cancer Scale
        - Decision Regret scores

2. Unconventional Data (Digital Phenotyping): this data will be collected continuously and unobtrusively from participants using their personal smartphones and a project-provided wearable device.

    a. Data from Smartphones:
        - Data gathered from a paired wearable device
        - Contextual smartphone usage data (e.g., call log, SMS log, app usage per category, device activity, WiFi connections)

    b. Data from Wearable Devices:
        - Physical Activity and Behavioral Indicators: Step counter, accelerometer, gyro sensor, daily activity levels, mobility indices, and sedentary patterns
        - Daily Activities and Isolation: GPS tracker (to be translated to points of interest and subsequently to activities)
        - Biomedical Parameters and Vital Signs: Sleep duration and stage, blood oxygen, blood pressure, heart rate and variability, body composition (from bioelectrical impedance, electrical heart sensor, optical heart rate sensor, PPG sensor)
        - Context Data: Barometer, light sensor

NON-PERSONAL AND STAKEHOLDERS' DATA:

Non-Personal Data: This category includes all the data sets that do not relate to an identifiable person. In the LATE-AYA project, this data is primarily generated as a result of research and project management activities.

1. Aggregated Datasets and Models: The digital phenotyping will support the creation of the core dataset “AYA cancer survivor dataset” (D7.1). While the data is highly personal, the project will only use it to develop aggregate statistical and AI models (D7.2). These models are considered non-personal data resources

[1] The specific list of questionnaires is still being determined

(especially when anonymized and aggregated). The ambition is to disseminate the models via platforms such as UNCAN.

2. Research Outputs: the project will produce numerous reports, plans, and guidelines, which are non-personal. Examples are the "Co-creation report (D2.2)", "Intervention report (D8.1)", policy briefs, and cluster meetings conclusions.

3. Cloud Services and Technical Data: The project that uses cloud services for data storage. This refers to technical and operational data, such as platform/service metadata, system performance and security logs, access logs for the EHDS-aligned SPE, data pipeline telemetry, data quality rule execution reports, and anonymised usage statistics needed to operate the SPE and analytics stack.

Stakeholders' Data: the project is highly collaborative and involves many stakeholders in addition to the AYA participants. Data gathered from these sources (mainly through co-creation activities) will inform the project.

1. Co-creation and User Needs Data: WP2 is dedicated to "Co-creation" and aims to convene all key project stakeholders (HCPs, caregivers, researchers, ethicists, innovators). The project will employ methods such as focus groups, interviews, workshops, polls, and questionnaires to obtain data from the said groups regarding needs mapping and feedback collection. Data collected through questionnaires and/or focus groups will be anonymous and will not contain any sensitive data. The deliverables, like "Detailed user needs specification" (D2.1), are based on the gathered stakeholder data.

2. Data from clinical staff: This dataset is not explicitly elaborated in the study, but it is presumably collected from clinical staff at the seven hospitals. This consists of expert opinions in the study protocol design, data quality rules, and feedback on digital tools.

3. IPR and Business Planning Data: The project will collect data from partners on background IPR and exploitation/business intentions (WP10). This is data from project stakeholders, namely the beneficiaries.

4. Perceptions on Ethical, Legal, and Social Issues (ELSI): Qualitative data (interviews and narrative answers in TXT format) obtained from experts related to clinical studies, AI use, and data protection.

RE-USE OF EXISTING DATA:
The project plans to re-use the following categories of existing data to establish context, train AI models, and ensure compatibility with current medical practices:

1. Retrospective/Historical Clinical Data: Collected indirectly from hospitals and research institutions through their information systems and Electronic Health

Records (EHRs). This data provides the clinical and demographic context necessary to identify AYA cancer survivor subgroups, identify risks for late effects (LE), and establish a benchmark for evaluating intervention tools.

2. Data from Epidemiological Studies and Registries: These external datasets may be repurposed to help measure the efficacy of the digital intervention developed by LATE-AYA.

3. Existing Behavioral Datasets: Datasets from similar health domains may be considered for use with transfer learning techniques to support the development of AI models.

In summary, while the primary dataset in the LATE-AYA project entails sensitive personal data from AYA cancer survivors, the project also collects non-personal data (aggregated models and research outputs) and extensive data about its multifaceted stakeholders to steer the co-creation process and ensure the relevance of its outcomes.

b. Purposes of Data Collection and Generation Within LATE-AYA

The collection and processing of data in the LATE-AYA project are central to achieving its objectives. The project addresses the fragmentation between clinical records and daily life data of adolescent and young adult (AYA) cancer survivors, both of which are needed to understand and manage the late effects of cancer treatment. The project's primary objective is to develop a secure, privacy-preserving, and patient-centered digital ecosystem for personalized follow-up, which focuses on preventive health behaviors, psychological support, and social reintegration. This ecosystem will use advanced technology, including Artificial Intelligence.

Data is the core component for LATE-AYA, facilitating personalized care and scientific discovery. Table 1 explains why each category of data is collected or reused.

DATA TYPE and ACTIVITY	PURPOSE and RELATION TO PROJECT OBJECTIVES
Generation of Conventional Data (Clinical records, demographics, psychosocial PROMs)	This data is collected during clinical follow-up to provide full access to patient medical histories. This enables healthcare professionals to track health evolution and make necessary adaptations in time to anticipate late complications and other disease-specific aspects that impact quality of life
Generation of Unconventional Data (Digital phenotyping via smartphones/wearable devices: heart rate, sleep, activity)	This continuous data is used for AI-based digital phenotyping to quantify an individual's phenotype in real time. This enables the modeling of health trajectories, the prediction of LE risk, and the discovery of new digital biomarkers.

Re-use of Retrospective Clinical Data (Historical EHR data from hospitals)	This traditional data is necessary to create a baseline, provide the clinical and demographic context of the samples, and identify AYA cancer survivor subgroups and their risks for LE. This also helps establish a benchmark for evaluating the intervention tools.
Re-use of Epidemiological/Behavioral Datasets	External epidemiological studies and registries may be used to help measure the efficacy of the digital intervention. Existing behavioral datasets may be used with transfer learning techniques to support the AI models.
Generation of Stakeholders' Data (Co-creation, interviews, workshops, surveys from clinicians, caregivers, ethicists, etc.)	This data informs the "Detailed user needs specification" (D2.1), helping the consortium collaboratively design and develop a safe, privacy-sensitive digital space. This process ensures the digital tools reflect survivors' needs and integrate into healthcare practice.
Generation of Clinical Trial Data (Two-stage trial involving 400–450 AYA survivors)	The clinical trial data (Stage 1 observational, Stage 2 randomized) is used to generate evidence and assess the effectiveness of the digital intervention. This directly supports the project's aim of delivering evidence-based solutions for cancer survivorship.
Generation of Technical Data (System logs, performance metrics, software code)	This data ensures the proper functioning and performance of the digital toolkit and the SPE. For example, a partner's software code is foreground IPR used to support the project's technological objectives, such as automating personalized health messages through an AI-based LLM.

Table 1: Data, activities and relation to project objectives

Additionally, all data collection and generation activities are undertaken with explicit consideration of interoperability standardisation, integration with European research infrastructures, and the creation of durable public assets.

c. Data Format and its Size

The project targets a sizeable data scale because its multiple-stage clinical study will monitor about 400–450 AYA cancer survivors aged 16 to 39[2] from three European Union countries. The study will enrol at least 100 AYAs beyond cancer in observational Stage 1, and approximately 300–350 participants in randomized Stage 2, who will be followed for 18 to 30 months. The project collects large amounts of both conventional and unconventional data, with the unconventional data being collected on a continuous

[2] The viability of having underage participants is still being discussed within the consortium given the need to ensure the scientific validation of the applicable questionnaires for the target age ranges.

moment-by-moment basis through continuous monitoring data streams from smartphones and wearables.

The data collected and reused exists in a variety of formats. Unconventional data, such as the physiological and behavioral signals collected continuously from wearable devices and smartphones, is primarily stored in machine-readable open formats like JSON and CSV. Additionally, data gathered through questionnaires on PROMs often exists in structured formats such as Digital spreadsheets (.xlsx), CSV, and SQL. Additionally, data derived from co-creation activities and expert interviews, such as raw narrative answers, are typically stored in TXT format, and workshop materials may include meeting recordings in MP4 format. Furthermore, technical data, including software components, utilizes machine-readable formats such as CSV, JSON, and XML.

Lastly, all formats of the device and clinical data resulting from this project will be converted into the established standard to ensure interoperability (reusability) in line with the FAIR principles. The two mandatory interoperability standards for FDA submissions are OMOP (Observational Medical Outcomes Partnership) and FHIR (Fast Healthcare Interoperability Resources). This is necessary to ensure that the project datasets and AI models can be reused and integrated in a clean way within the European infrastructures (e.g., UNCAN).

d. Identified Datasets within LATE-AYA

Table 2 summarized the identified datasets within LATE-AYA.

DATASET WORKING NAME (PARTNER)	CONTENT/ DESCRIPTION	PRIMARY FORMATS/ STORAGE	PID STATUS	ACCESS/ RESTRICTIONS	RETENTION PERIOD	LICENSE/ PUBLIC AVAILABILITY
Late-AYA Digital Phenotyping Datatset (DOT)	Time-series physiological and behavioral signals from wearable devices and smartphones, linked to pseudonymised IDs. Used for exploratory ML models	JSON, CSV; Encrypted at rest (AES-256). Stored in Secure CERTH cloud.	DOIs will be assigned at dataset release.	Restricted; limited to authorized project partners. Anonymised dataset may be deposited in Zenodo.	Project duration + 5 years (for non-anonymized data)	Anonymised datasets under CC-BY-NC 4.0 (future reuse)
Clinical Data (VHIO/VHIR)	Clinical, treatment-related, and Late-Effect (LE) data; includes ePRO and PROM questionnaires (e.g., EORTC, PROMIS measures). Data is pseudoanonymised	Excel, CSV, SQL. Stored in REDCap.	A DOI will be assigned	Controlled access through Data Access Committee (DAC) and Data Access Agreement (DAA).	Information not available at this stage of the project	Data will be anonymized or aggregated prior to publication.

DATASET WORKING NAME (PARTNER)	CONTENT/ DESCRIPTION	PRIMARY FORMATS/ STORAGE	PID STATUS	ACCESS/ RESTRICTIONS	RETENTION PERIOD	LICENSE/ PUBLIC AVAILABILITY
Population_ Data[3] (INT)	Clinical, demographic, lifestyle, and psychosocial data for Stage 1/2 trials, forming the foundation for AI/ML algorithm development.	REDCap environment. Clinical data follows HL7 FHIR format.	Unique and persistent DOI registered through Zenodo will be assigned.	Restricted access subject to institutional ethics committee approval and formal sharing agreement.	10 years following project conclusion.	Open datasets under Creative Commons Attribution license (CC BY 4.0).
patients_ieo (IEO)	Structured dataset including clinical history, device-generated data (wearable devices), PROMs, and psychosocial assessments used to train/validate the AI model.	Structured eCRF REDCap data, exported in CSV. Stored in a secure cloud environment	Plan to use a recognized and repositorys upported persistent identifier (DOI or PURL).	Strictly restricted; not publicly available due to sensitive clinical information. Access requires formal data-sharing agreement and ethics approval.	10 years after project conclusion.	Aggregated study results released under CC BY 4.0.
Clinical/Wea rable Data (NVI)	Clinical data (demographics, cancer type) and wearable device data	CSV. Stored on secure servers.	Information not available at this stage of the project	Restricted; Consortium use only. Access protected by authorization/aut hentication.	5 years in the electronic database	Under Data Sharing agreement between the consortium.
StageI_data set + StageII_data set (UPM)	Clinical information, demographics, treatment history, QoL, and LE-specific questionnaires from AYA survivors.	Not yet defined (likely JSON and/or CSV). Stored within the Secure Processing Environment (SPE).	Expected to be discoverabl e via the secure processing environmen t.	Access by LATE-AYA partners. Expected to share for re-use in EU infrastructures (UNCAN.eu and EHDS).	Probably about 5 years after project ends - Information not clear at this stage of the project	Information not available at this stage of the project
Perceptions on ELSI (UGR)	Qualitative statements from interviews with experts regarding ethical, legal, and social issues (ELSI)	TXT or CSV. Stored in UGRDrive	Information not available at this stage of the project	Restricted access for raw data; aggregated data open access. Deposited in OSF or DIGIBUG.	Information not available at this stage of the project	CC BY 4.0
Stakeholder Requiremen ts Dataset (UTW, CQ, SAM)	Stakeholder requirements, user stories, qualitative comments for digital service design (WP2). SAM includes raw Wearable/Clinical Data.	.xlsx, .csv, .txt, .docx. Stored on institutional servers (UTW/CQ) or Amazon S3 Bucket in Europe (SAM).	No persistent identifiers are planned.	Internal consortium use only. Not shared publicly.	UTW/CQ: 10 years. SAM: 5 years.	Not applicable (Internal requirement documents).
LATE-AYA Co-Creation	Data collected during WP2 workshops and co-	Digital text documents (.docx, .txt),	No persistent identifiers	Restricted access for WP2 partners and authorized	Information not available at this	Not Applicable

[3] Ongoing discussions surround the viability l to have a single centered patient dataset potentially stored a CNAO.

DATASET WORKING NAME (PARTNER)	CONTENT/ DESCRIPTION	PRIMARY FORMATS/ STORAGE	PID STATUS	ACCESS/ RESTRICTIONS	RETENTION PERIOD	LICENSE/ PUBLIC AVAILABILITY
Workshop Dataset (UTW)	creation sessions with stakeholders (clinicians, developers, patient representatives, caregivers). Includes notes, audio transcripts, whiteboard inputs, and anonymized summaries of discussions on functional and non-functional requirements for LATE-AYA components.	Excel sheets (.xlsx), and meeting recordings (.mp4).	are planned.	consortium personnel. Internal consortium use only. Data anonymized before cross-WP sharing. Not for external or public dissemination.	stage of the project	
HTA for impact (KTU)	Primary and secondary patients and other stakeholders, like healthcare professionals, experience using the LATE AYA digital system data. Datasets including Population data to early assess the clinical impact of the LATE AYA digital system.	Pseudonymised individual-level data in machine-readable formats (e.g., CSV, XLSX, SQL databases, JSON); survey exports (CSV/XLSX), system log files, and documentation/metadata in PDF/DOCX and README (TXT/MD). Aggregated tables and analysis files (e.g., R / Stata / SPSS / Python project files).	Typically DOIs, will be assigned to sharable datasets upon release in trusted repositories . Metadata will be rich and machine-readable. Project-internal data will be catalogued with unique internal identifiers.	Restriczed access to raw data only avaialble to team members at KTU and partner institutions under confidentiality. Anonymised/ aggregated datasets underpinning publications: may be shared via a trusted repository under controlled or open conditions if re-identification risk is negligible and ethics/funder requirements allow.	At least 5 years after the end of project, in line with EU/KTU open access and RDM guidance; longer retention may apply where required by medical data legislation, good research practice, or funder/publisher requirements (up to ~10 years or as specified in institutional policy).	No open licence for identifiable or pseudonymised patient-level data (access governed by data processing and data-sharing agreements).
Population_ Data (UNTO)	Dataset including data on the AYA population which will be built in two phases: the first corresponding to the observational, stage 1, clinical trial, the second corresponding to the interventional,	REDCap environment	Unique and persistent Digital Object Identifiers will be assigned to each dataset release (registered through	Access granted to authorized personnel only. Access to this data will be subject to approval by the institutional ethics committee and contingent upon the establishment of	Datasets will be made reusable for 10 years following the conclusion of the project.	Not provided

DATASET WORKING NAME (PARTNER)	CONTENT/ DESCRIPTION	PRIMARY FORMATS/ STORAGE	PID STATUS	ACCESS/ RESTRICTIONS	RETENTION PERIOD	LICENSE/ PUBLIC AVAILABILITY
	randomized, stage 2, clinical trial		Zenodo) to guarantee long-term citation and version stability.	a formal data sharing agreement that outlines the intended use, safeguards, and data handling responsibilities of the requestor.		

Table 2: Identified datasets within LATE-AYA

e. Data Utility Outside of the Project

The LATE-AYA project data and assets use beyond the project (data sharing and standardization policy, clinical implementation, and commercial exploitation). A critical pathway is the implementation of project assets into the European Health Data Ecosystem. The SPE will be EHDS-ready and adhere to the FAIR principles and models obtained during the project implementation. To this aim, the consortium is planning to list its large multisource datasets with relevant Health Data Access Bodies (HDABs) according to EHDS Regulation article 36 within three years after the end of the funded project.

The outputs are highly important for scientific research and policy-making. Researchers will gain access to a unique dataset (D7.1: AYA cancer survivor dataset) that combines conventional clinical data with novel digital phenotyping data. This resource enables the investigation of advanced Statistical and AI models (D7.2) for predicting Late Effects (LE) onset and modelling HRQoL trajectories. These findings are directly translated into practice via dedicated deliverables, including Recommendations for inclusion into clinical guidelines (D9.3) and annual Policy Briefs (D10.14 - D10.18) targeting regulators and national ministries.

The project's planned utility also includes robust pathways for commercial exploitation and long-term sustainability, led by industry partners. Work Package 10 (WP10), led by SAMSUNG ELECTRONICS (UK) LIMITED (SAM), is responsible for defining the strategy to ensure the LATE-AYA tools are reused in future research endeavors and innovation initiatives, after the end of the project. This involves identifying components (developed in WP3-WP5) for potential commercial exploitation (of SW components, knowledge assets, or services). This market potential is formalized in the Business Plan (D10.2), which documents measures to exploit project results scientifically, socially, and commercially. Furthermore, an Early Health Technology Assessment (HTA) (D9.1) is conducted to preliminarily assess the cost-effectiveness of the LATE-AYA solutions compared to the standard of care, providing the necessary economic justification for adoption by healthcare payers.

This project has received funding from the European Union's EU4Health Programme under Grant Agreement No 101214326

2. Fair Data

The European Commission requires the application of FAIR principles (Findable, Accessible, Interoperable, and Reusable) to research data in Horizon Europe. In the LATE-AYA project, these principles will govern how data is described, preserved, and disseminated to optimize its scientific and societal value.

a. Findability

According to the principle of findability, the data needs to be found. "Thus, metadata and data should be easy to find for both humans and computers."[4] The LATE-AYA project takes a varied but often strict approach to ensuring the findability of its different research outputs. It makes clear commitments to use persistent identifiers and structured metadata for the most important datasets. However, it pays less attention to external standards for internal and technical data streams.

When it comes to persistent identifiers (PIDs), the approach varies across the different types of data generated. Major clinical and digital phenotyping datasets will have a Persistent Identifier (DOI). For instance, the LATE-AYA Digital Phenotyping Dataset and the clinical data managed by VHIO/VHIR will receive a DOI when they are released. On the other hand, for other significant datasets, like the Perceptions on Ethical, Legal, and Social Issues (ELSI) dataset, the plan for PIDs is still pending. Additionally, for technical and internal outputs, such as software code and configuration files managed by Vidyasoft, there are no plans for persistent identifiers. The internal stakeholder requirement datasets from the University of Twente (UTW) will also not include persistent identifiers, as they will not be made public.

To ensure general reusability, all datasets will be supported by documentation. This includes README files that describe the dataset structure and content, as well as codebooks with variable definitions. Software releases will also include README files, data dictionaries, and technical manuals that explain structure, functionality, and usage. Provenance will be recorded through standardized metadata. This will detail the origin, version history, processing steps, and parties responsible.

The LATE-AYA project enforces metadata standards for interoperability and discoverability, utilizing two main categories of standards:

1. General/Open Standards: The Digital Phenotyping Dataset uses JSON-LD metadata following the Dublin Core and DataCite schemas.
2. Clinical Standards: Crucial clinical and wearable data are mapped to established standards such as OMOP (Observational Medical Outcomes Partnership)[5] and FHIR (Fast Healthcare Interoperability Resources) to ensure compatibility within

[4] GO FAIR Initiative. *FAIR principles.* https://www.go-fair.org/fair-principles/
[5] To be determined given compatibility with questionnaires.

the Secure Processing Environment (SPE) and with EU infrastructures. Specific clinical partners intend to use widely accepted medical vocabularies like MedDRA and CTCAE.

For all open project results, search keywords will be provided in the metadata. It should be possible to harvest and index the data through open-access collective repositories and schemas. The metadata of the ELSI dataset will be registered on platforms such as OSF or DIGIBUG to be indexed by search engines. The main clinical data repositories will use the Dublin Core and DataCite metadata schema. Main clinical data repositories will be set up in compliance with the FAIR principles (findable, accessible, interoperable, and reusable), based on standards such as Dublin Core or DataCite to enable machine harvesting and indexing by external services. Conversely, the metadata of internal restricted datasets (e.g., UTW or Vidyasoft) will only be stored in internal repositories and documentation systems.

b. Accessibility

According to the principle of accessibility, after the data are found, users need to know how they can be accessed, possibly including authentication and authorisation. [6] Throughout LATE-AYA, accessibility will be assured by providing clear usage conditions that balance open science goals with GDPR rules.

REPOSITORY:
Non-confidential datasets from LATE-AYA partners are generally planned for deposition in trusted repositories like Zenodo or OSF/DIGIBUG. Nevertheless, the large, invaluable datasets collected, including both "conventional clinical data" and "unconventional patient-generated data", will all together be securely stored in a privacy-preserving cloud repository. This repository is the SPE, which will be designed and implemented by CERTH. The SPE is intended to be the central storage location for the datasets generated during the clinical pilot study. In the later stages of the project, digital assets, including datasets and predictive models, will be deposited and integrated into the UNCAN.eu platform. This platform is a research data platform based on a network of National Cancer Data Nodes, contributing to the broader development of a robust and sustainable European Open Science Cloud (EOSC). Moreover, the consortium is also planning for integration with the formal EU data sharing ecosystem. They plan to catalogue the LATE-AYA datasets with relevant Health Data Access Bodies (HDABs), as per EHDS Regulation Article 36, within three years after the end of the funded project.

At the current stage of the project, the following partners have indicated that they are planning on assigning persistent identifiers: INT, DOT, VHIO and VHIR, IEO. The rest of the partners are not using identifiers as they are either not planning on it (UTW, CQ, VS) or are not sure about it at this stage of the project (UGR or UDEU).

[6] Ibid

DATA

Not all data collected and reused within the project will be made openly available. Public access is largely restricted due to the collection of Special categories of personal data (health, psychosocial, and digital phenotyping data), requiring strict adherence to the GDPR. Intentional restrictions apply to raw data and internal documentation, which are for consortium use only (e.g., Stakeholder Requirement Dataset). However, open sharing is reserved for anonymised and aggregated datasets (e.g., IEO). Software will also be accessible; nevertheless, it may undergo a short embargo period for final testing and validation after project completion (VS).

Data of partners like DOT and IEO will be available through free and standard access protocols like HTTPS and secure REST APIs (e.g., DOT, IEO). Nevertheless, some partners with a lowered level of access will impose verified authorization and authentication methods such as OAuth2 (DOT) or institutional accounts (UTW). In case of partners like VHIO and VHIR a Data Access Committee is needed to review and approve access requests for personal or sensitive data.

METADATA

According to the Grant Agreement, the data should be kept available and findable beyond the end of the project, until May 31, 2030. The duration is linked to the exploitation strategy and commitment to regulatory authorities in the EU. Moreover, grant beneficiaries must make every effort to exploit their results (including data and models) for up to four years following the project's conclusion. Additionally, the metadata is guaranteed to remain available and open even if the underlying data itself is no longer accessible or usable, as per the Open Science requirements of the Grant Agreement.

Documentation about the necessary software will be included. All software releases will have README files, data dictionaries, and technical manuals that explain their structure, functionality, and usage. The relevant software, including code and configuration files, may be made publicly available after the project ends through open repositories. While developers will use permissive open-source libraries, such as Python, NumPy, and React, the software is initially limited by an internal consortium license agreement.

c. Interoperability

According to the principle of interoperability, data usually needs to be integrated with other datasets and must also interoperate with applications or workflows for analysis, storage, and processing.[7] Throughout LATE-AYA, interoperability will be achieved by using established standards and terminologies, including OMOP and FHIR (UPM, INT, DOT).

Beyond OMOP and FHIR, the project incorporates other standards and shared

[7] Ibid

vocabularies to maximize interoperability. For example, IEO will adopt shared vocabularies and ontologies for clinical and questionnaire data to ensure semantic consistency and interoperability. Secondly, NVI aims for interoperability by using standards such as MedDRA and CTCAE. Additionally, VS will adopt standardized metadata schemas and controlled vocabularies for interoperability, and they will also use existing domain-specific vocabularies where applicable to ensure semantic interoperability. Lastly, VHIO and VHIR will ensure interoperability by using standard, widely accepted data and metadata formats, and they plan to follow Community-endorsed vocabularies and ontologies whenever available.

The most common machine-readable formats across the consortium are CSV and JSON. Clinical and questionnaire data that have been collected in REDCap are typically exported in CSV format, which is the format also sometimes used for the phenotyping data. JSON is primarily used for the digital phenotyping dataset and software-related data. Partners also frequently use XML, SQL, TXT, DOCX, and XLSX files. Together, these open formats ensure that the consortium can store, share, and reuse diverse types of data.

d. Reusability

According to the principle of reusability, the ultimate goal of FAIR is to optimise the reuse of data, which requires that metadata and data are well-described so they can be replicated and combined in different settings.[8] The LATE-AYA project ensures reusability by appropriately documenting the clinical and digital phenotyping data via README files, codebooks, and data cleaning procedures.

Due to the sensitive nature of the data, it is not publicly available. Access is subject to a Data Access Committee's (DAC) approval and a Data Access Agreement (DAA), and it complies with the European Union General Data Protection Regulation (EU GDPR). The anonymised outputs (i.e., the Digital Phenotyping Dataset) will be deposited in public repositories under licenses such as CC-BY-NC 4.0 or CC BY 4. More specifically, data will be usable by third parties post-project, especially via EU infrastructures like UNCAN.eu and EHDS. Provenance of the data will be thoroughly documented using standardized metadata detailing origin and version history.

Lastly, the project focuses on research outputs beyond just data, like software and models. It allocates resources to data security with methods like pseudonymisation at the source, AES-256 encryption, and secure cloud environments. Ethical compliance involves informed consent, specific ethical approvals such as those from the UTEW BMS Ethics Committee and clinical IRB before data collection, and set data retention periods of 5 years or 10 years.

[8] Ibid

3. Other Research Outputs

MANAGEMENT OF DIGITAL OUTPUTS

The LATE-AYA project adheres to the principle that outputs such as software, models, and protocols must be managed and shared in line with FAIR principles. The key digital research outputs managed under FAIR principles include:

1. Software and Algorithms (Foreground IPR)
2. AI Models (e.g., LLM and Exploratory ML Models)
3. Project Documentation and Qualitative Datasets (e.g., Co-creation outputs)

Table 3 explains measures implemented throughout the project to ensure the application of FAIR principles to the digital outputs.

FAIR PRINCIPLES	IMPLEMENTED MEASURES
Findability	• Several of the partners will assign Persistent Identifiers (DOIs) at the dataset release for outputs like the digital phenotyping dataset and clinical datasets. • It is planned to use the standard metadata formats, including custom JSON-based schemas for software to ensure discoverability. DOTSOFT specifically intends to use JSON-LD metadata following the Dublin Core and DataCite schemas.
Accessibility	• Software and related assets are planned for long-term preservation and sharing in Open repositories such as Bitbucket or the consortium's institutional repositories. Anonymised datasets may be deposited in open-access repositories like Zenodo. • Data and software will be accessible via HTTPS and APIs where appropriate. Internal access to datasets, such as transcribed ELSI interviews, will be provided through the consortium's SharePoint. • The AI and software developed by partners like Vidyasoft may be made publicly available after project completion, possibly after a short embargo period.
Interoperability	• Software-related data and components are stored in machine-readable formats such as CSV, JSON, and XML. Qualitative outputs are made available in TXT or CSV formats. • Interoperability for clinical and wearable datasets is ensured by mapping the data and

FAIR PRINCIPLES	IMPLEMENTED MEASURES
	metadata to standards like OMOP (Observational Medical Outcomes Partnership) and FHIR (Fast Healthcare Interoperability Resources). • Community-endorsed vocabularies and ontologies will be followed whenever available. For clinical data, standards like MedDRA and CTCAE are intended for use.
Reusability	• Datasets and software releases will be accompanied by detailed documentation to support validation and reuse. This documentation includes README files, data dictionaries, and technical manuals explaining the structure, functionality, and usage of the software. • For general research outputs, licenses such as CC BY 4.0 are planned. Anonymised digital phenotyping datasets may be released under a CC-BY-NC 4.0 licence for future FAIR reuse. Software components, initially governed by an internal consortium licence agreement, are planned for public release under permissive licensing after the project concludes

Table 3: Measures to implement the FAIR principles

MANAGEMENT OF INTELLECTUAL PROPERTY RIGHTS

The Intellectual Property Rights (IPR) for research outputs, including software and technical know-how, are managed through the Consortium Agreement, ensuring compliance with Horizon Europe requirements. A dedicated task in the project will carry out the mapping and management of IPR to ensure their proper protection, the validation of sustainability of project outputs and general alignment with the exploitation activities.

4. Allocation of Resources and Data Management Lead

As task manager and project coordinator, Universidad Politécnica de Madrid (UPM), is tasked with leading T1.4 and addressing quality, risk and data management oversight (including the data management coordination activities of the AYA cluster, the initial results of which can be found in Annex 1 of this deliverable). This being said, Data Management activities will be coordinated with the support of UDGA, which has been integrated into T1.4 as the specific Lead Beneficiary responsible for creating and submitting the official DMP deliverables (D1.4 and D1.5).

The financial architecture underpinning the project's data management is supported through a combination of eligible EU grant expenditures and significant external funding. The direct EU grant allocates funds for critical data infrastructure, including €50,000 specifically designated as a purchase cost to CERTH for Cloud services for data storage, necessary for establishing the Secure Processing Environment (SPE). Additionally, an eligible cost of €200,000 is budgeted for SAM under equipment costs for the wearable devices used in clinical studies, which are essential for generating the vast digital phenotyping datasets. However, the specialized legal and data expertise required for drafting the Data Management Plan (D1.4 and D1.5) and ensuring compliance with regulations like GDPR and EHDS is contributed by the Associated Partner UDG Alliance (UDGA) funded externally by SERI and are not charged to the EU grant

No additional costs are expected at this stage since all partners have person-months allocated to WP1 and are jointly responsible for this activity as agreed in the Grant Agreement. Any relevant updates regarding the resource requirements or costs will be included in subsequent versions of the DMP.

5. Data Security, Storage, and Retention

Data collected within the LATE-AYA project will be managed through a centralized architecture to ensure robust security, interoperability, and FAIR compliance. The SPE will be cloud-based and will cover all the datasets collected under the clinical study, including personal data collected using mobile and wearable devices that may include sensitive data. This central repository is built on the principles of Zero Trust for maximum data security and is compatible with the future needs of the European Health Data Space (EHDS) Regulation. Personal data processing shall be done in a dedicated data analytics and ML/AI platform over the centralized SPE. Access-controlled, secure platform will allow researchers to search, train, and validate models against pseudonymized data from edge devices and clinical sites.

For purposes of guaranteeing security, technical procedures employed throughout the duration of the project will include advanced encryption, anonymisation, and pseudonymisation techniques. In addition to this, role-based access controls with a separate authorisation framework will control different levels of access. Management of the data will also include system monitoring and audit, periodic checks on data quality, and risk assessments in tandem with a backup plan.

The project intends to use specific repositories (OSF, Zenodo, Bitbucket, Github) for long-term preservation and data sharing, particularly for anonymized and aggregated results. Most partners agreed on storing the documents and data for 5 to 10 years. Beneficiaries must maintain records and supporting documents for five years from receipt of the final payment. At the completion of a project, new information in terms of digital datasets and ML/AI models will be released for reuse through facilities such as the UNCAN.eu platform, according to the EHDS Regulation on secondary use of health data. The overall data management plan will meet the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) so that this integration is possible.

This project has received funding from the European Union's EU4Health Programme under Grant Agreement No 101214326

6. Ethics and Regulatory Compliance within the LATE-AYA Project

The LATE-AYA project is committed to upholding the highest ethical standards and to compliance with the regulatory framework established under Horizon Europe, including the provisions of the Grant Agreement (Articles 14–15 on ethics and data protection). All project activities must be aligned with relevant EU legislation, including but not limited to the General Data Protection Regulation (GDPR), Digital Services Act (DSA), the Artificial Intelligence Act (AI Act), the CRA, and the NIS2 Directive.

In regard to the ethical side of the project, the European Code of Conduct for Research Integrity provides the foundational baseline guiding ethical and responsible research practices.[9] Additionally, the project is required to utilise instruments such as the Universal Declaration on Human Rights, the European Convention on Human Rights, and the Declaration of Helsinki, among others. Furthermore, ethical approval will be obtained from the relevant authorities before data collection, and continuous oversight will ensure that the rights, dignity, and well-being of AYA cancer survivors remain central throughout the project. The Late-Aya Regulatory Framework is thereby declared in the following sections as a baseline element for downstream activities which may need to connect with data-specific considerations.

a. LATE-AYA Regulatory Framework

General Data Protection Regulation

The General Data Protection Regulation (GDPR) lays down rules on the processing of personal data of individuals within the European Union,[10] regardless of the location of the data controller or processor, provided that the processing is related to the provision of goods or services to, or the monitoring of behaviour of, individuals within the EU.[11] Processing of personal data is defined as "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction."[12] The Regulation differentiates between data controllers, those who determine the purposes and means of processing,[13] and data processors, those who

[9] ALLEA (2023) The European Code of Conduct for Research Integrity – Revised Edition 2023. Berlin.
[10] European Parliament & Council of the European Union. (2016, April 27). Article 1. 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, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119, 1–88.
[11] Ibid, Art 3
[12] Ibid, Art 4(2)
[13] Ibid, Art 4(7)

process data on behalf of controllers.[14] Depending on their role in specific tasks, LATE-AYA partners may act as either controllers or processors. The categories of data processed within LATE-AYA are described in Section 3.1 of this deliverable. These include not only personal data but also special categories of data, such as health and genetic data, the processing of which is by default prohibited under Article 9 GDPR unless a relevant exception from paragraph 2 of the Article applies.

For the lawful processing of personal and sensitive data, partners must rely on the appropriate legal bases listed in Articles 6 and 9 GDPR. In LATE-AYA, the most relevant legal bases are:

- Article 6(1)(a): consent of the data subject;
- Article 6(1)(b): processing necessary for the performance of a contract;
- Article 6(1)(c): processing necessary for compliance with a legal obligation (e.g., EHDS requirements);
- Article 6(1)(f): processing necessary for legitimate interests, unless overridden by the rights and freedoms of the data subject, especially when children are involved;
- Article 9(2)(a): explicit consent of the data subject for processing special category data;[15]
- Article 9(2)(i): processing necessary for reasons of public interest in the area of public health;
- Article 9(2)(j): processing necessary for archiving in the public interest, scientific or historical research, or statistical purposes, in line with Article 89(1) safeguards.

Additionally to these, the GDPR establishes specific dispositions for processin gof personal data for scientifi research as part of its Article 89. Secondary use of personal data has been examined by the European Data Protection Board (EDPB) which considers that secondary use of personal data for scientific research can be treated as compatible with the original purpose under Article 5, provided that the controller implements the safeguards required by Article 89(1), including data minimisation, pseudonymisation (where possible), and organisational/technical measures.

Throughout the project, it is imperative to ensure that the rights of data subjects are upheld in accordance with Articles 12–22 of the GDPR. This encompasses the rights of access, rectification, erasure, restriction, objection, and data portability. Test participants in LATE-AYA must be fully informed of these rights and provided with clear means to exercise them.

Furthermore, partners are obligated to adhere to the principles of data processing

[14] Ibid, Art 4(8)

[15] On consent, the EDPB generally accepts broad consent (Recital 33) only within well-defined scientific domains and rejects open-ended future-use consent. It stresses that secondary research involving special-category data still requires its own lawful basis most often Article 9(2)(j) combined with Member-State-level research legislation and that any derogation from data-subject rights under Article 89(2) must be strictly necessary to avoid impairing the research objective.

stipulated in Article 5 of the GDPR, namely: lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, confidentiality, and accountability.

Further obligations are applicable under the following legislation:

- Article 25 stipulates the principles of data protection by design and by default.
- Article 32 stipulates the implementation of appropriate technical and organisational security measures. Such measures include encryption, pseudonymisation, access control, and incident monitoring.
- Articles 33–34 stipulate the reporting obligations that must be fulfilled in the event of a personal data breach. Such obligations include the notification of the relevant supervisory authorities and, where applicable, the affected individuals.
- Article 35 stipulates the necessity of conducting a Data Protection Impact Assessment (DPIA) before the undertaking of high-risk processing activities, which are defined as the handling of health or genetic data.

European Health Data Space Regulation

The European Health Data Space Regulation (EHDS) is a Regulation which provides common rules, standards, and infrastructures and a governance framework, with a view to facilitating access to electronic health data for the purposes of primary use of electronic health data and secondary use of those data, and establishes the European Health Data Space.[16] The primary use is defined as the processing of electronic health data for the provision of healthcare. This is to say, the technology is used for the assessment, maintenance, or restoration of one's health.[17] Secondary use is defined as the processing of data for purposes other than those for which it was originally collected or produced, provided that the conditions set out in Chapter IV are met.[18] Both purposes are relevant in the LATE-AYA project.

The EHDS differentiates between health data holders and health data users. A data holder is any legal or natural person who has the legal right or obligation to process personal electronic health data for purposes such as healthcare, research, innovation, or regulatory compliance; or can make available non-personal health data by controlling access, registration, or exchange mechanisms.[19] In LATE-AYA, beneficiaries such as HSJD or VHIR qualify as data holders. By contrast, data users are entities lawfully granted access for secondary use under a permit from national health data access bodies.[20] In practice, researchers in LATE-AYA accessing survivorship data for analysis would be considered data users. Health data holders and health data users are subject to distinct obligations, which are outlined in Table 4:

[16] European Parliament & Council of the European Union. (2025, February 11). Art 1(1). Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847. Official Journal of the European Union, L 2025/327, 1–120
[17] Ibid, Art 2(2)(d)
[18] Ibid, Art 2(2)(e)
[19] Ibid, Art 2(1)(t)
[20] Ibid, Art 2(1)(u)

ECONOMIC OPERATOR	OBLIGATIONS AND RESPONSIBILITIES
Health Data Holders	• **Art. 51:** Required to make minimum categories of electronic health data, such as EHRs and genomic data, available for secondary use. • **Art. 52:** Must inform the access body if requested data contains intellectual property or trade secrets and justify the need for protection • **Art. 60:** Must provide requested health data to a health data access body within three months and keep dataset descriptions in the national catalogue updated annually. • **Art. 78:** Must apply a Union data quality and utility label to datasets collected with public funding; this is optional for other datasets.
Health Data Users	• **Art. 54:** Strictly forbidden from using data for purposes detrimental to individuals, such as for insurance decisions, marketing, or developing harmful products. • **Art. 61:** Must only access data with a valid permit, not share it with unauthorized third parties, not re-identify individuals, and make results public within 18 months. • **Art. 62:** Must pay transparent and proportional fees for data access, with the option to withdraw the request after being informed of the estimated cost. • **Art. 67:** Must submit a detailed application to an access body, specifying the purpose, data needed, and planned safeguards. • **Art. 69:** May submit a health data request to receive a response in an anonymised statistical format without directly accessing the underlying data.

Table 4: EHDS stakeholder obligations and responsibilities

Other Articles that support the SPE design within LATE-AYA that should be taken into consideration are as follows:

- Article 66: safeguards on data minimisation and purpose limitation.
- Article 71: individual right to opt out of secondary use.
- Article 73: use of secure processing environments for access.
- Sections 4 & 5: rules on cross-border infrastructure, and quality/utility standards for secondary use of health data.

Artificial Intelligence Act

The Artificial Intelligence Act (AI Act) represents the European Union's first

comprehensive legislative framework governing artificial intelligence systems placed on or made available within the internal market. Its central aim is to advance human-centric and trustworthy AI by safeguarding health, safety, and fundamental rights enshrined in the EU Charter, while mitigating potential risks and adverse impacts of AI systems, and at the same time encouraging innovation.[21]

The AI Act differentiates AI systems by assigning them to four tiers of regulatory concern: unacceptable-risk systems, which are prohibited;[22] high-risk systems, which are subject to extensive compliance obligations;[23] limited-risk systems, which must meet specific transparency requirements;[24] and minimal-risk systems, which remain largely outside the scope of regulation. It also differentiates between different types of AI operators, such as providers, product manufacturers, deployers, authorised representatives, importers, or distributors.[25] Depending on the classification of the AI systems, different obligations for AI operators apply (Table 5).

RISK LEVEL	OPERATORS	OBLIGATIONS AND RESPONSIBILITIES
Unacceptable	**All**	• **Article 5:** Must not place on the market, put into service, or use AI systems that fall under the prohibited practices list. • **Article 99(3):** Non-compliance is subject to the highest level of administrative fines.
High-Risk	**Providers**	• **Article 16:** Fulfill comprehensive obligations, including ensuring compliance with all requirements, having a quality management system, drawing up technical documentation, and performing conformity assessments. • **Articles 9-15:** Ensure the system meets requirements for risk management, data governance, human oversight, transparency, accuracy, and cybersecurity. • **Article 47:** Draw up an EU declaration of conformity.

[21] European Parliament and of the Council. (2024, June 13) Art 1(1). Regulation (EU) 2024/1684 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L 22024/1689, 12.7.2024

[22] Ibid, Art 5

[23] Ibid, Art 6

[24] Ibid, Art 50

[25] Ibid, Art 3(8).

This project has received funding from the European Union's EU4Health Programme under Grant Agreement No 101214326

RISK LEVEL	OPERATORS	OBLIGATIONS AND RESPONSIBILITIES
		• **Article 48:** Affix the CE marking. • **Article 49(1):** Register the system in the EU database. • **Article 72:** Establish a post-market monitoring system.
	Importers + Distributors	• **Article 23 (Importers) + Article 24 (Distributors):** Act as gatekeepers by verifying that the high-risk AI system has the required CE marking, documentation, and declaration of conformity before making it available on the market
	Authorised Represetatives	• **Article 22:** Must be appointed by providers outside the EU. They are mandated to verify documentation, keep it available for authorities, and cooperate with them on the provider's behalf.
	Deployers	• **Article 26:** Use the system according to instructions, assign human oversight, monitor its operation, and keep logs. • **Article 27:** Certain deployers (e.g., public bodies) must conduct a fundamental rights impact assessment before use. • **Article 86:** Provide an explanation to affected persons for certain decisions based on the system's output.
General-Purpose AI	**Providers**	• **Article 53:** Draw up technical documentation, provide information to downstream providers, create a policy to comply with EU copyright law, and publish a summary of training content.
General-Purpose AI Models with Systemic Risk	**Providers**	• **Article 55**: Fulfill additional, stricter obligations, including model evaluations, assessing and mitigating systemic risks, tracking and reporting serious incidents, and ensuring a high level of cybersecurity protection.

RISK LEVEL	OPERATORS	OBLIGATIONS AND RESPONSIBILITIES
Limited Risk	**Providers + Deployers**	• **Article 50:** Ensure transparency for AI systems that interact with humans (chatbots), generate "deep fakes," or use biometric categorisation/emotion recognition. The key obligation is to clearly inform people they are interacting with an AI or viewing artificially generated content.

Table 5: AI act stakeholder obligations and responsibilities

The LATE-AYA project will develop an AI-driven digital phenotyping platform to manage the late effects of cancer treatment in young adult survivors. This platform will use Artificial Intelligence and Machine Learning models to analyze large datasets from smartphones and wearables to model health trajectories, predict risks, and discover new digital biomarkers. The project will also implement an AI-powered chatbot, using models like GPT and BERT, to deliver personalized psychosocial support and interventions through a conversational interface. From an initial evaluation of the AI system and its tasks, it should not be classified as a general-purpose AI model, nor a prohibited, high-risk, or limited-risk AI system. Nevertheless, each operator participating in the project bears the responsibility to determine, through a structured legal and technical assessment, whether and to what extent the provisions of the Artificial Intelligence Act apply to the systems they are developing or placing on the market.

Where the AI Act is determined not to apply, AI operators should nonetheless seek to align their development and deployment practices with the overarching principles of trustworthy AI. In order to do so, they should adhere to relevant soft law instruments, such as the Ethics Guidelines for Trustworthy AI.[26]

Medical Device Regulation

The Medical Device Regulation (MDR) establishes the rules governing the placing on the market, availability, and commissioning of medical devices and their accessories intended for human use within the European Union. It further extends to cover clinical investigations involving such devices and accessories when carried out in the Union.[27] Under this Regulation, the term “medical device” is defined as “any instrument, apparatus, appliance, software, implant, reagent, material or other article intended by the manufacturer to be used, alone or in combination, for human beings for one or

[26] High-Level Expert Group on Artificial Intelligence. (2019, April 8). Page 2. Ethics Guidelines for Trustworthy AI. Publications Office of the European Union. Retrieved from https://ec.europa.eu/digital-strategy/en/library/ethics-guidelines-trustworthy-ai

[27] European Parliament, & Council of the European Union. (2017). Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, Art. 1 (1), OJ L 117, 1–175. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745

more of the following specific medical purposes: diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease, (...)."[28] Within the LATE-AYA project, several components, such as the multi-function chatbot, follow-up dashboards, and virtual assistants, may fall within the scope of this definition. Alternatively, depending on how the individual elements of the LATE-AYA digital toolkit are characterised and ultimately placed on the market, they could also be classified as "accessories to a medical device".[29] The "accessory for a medical device" is defined as "article which, whilst not being itself a medical device, is intended by its manufacturer to be used together with one or several particular medical device(s) to specifically enable the medical device(s) to be used in accordance with its/their intended purpose(s) or to specifically and directly assist the medical functionality of the medical device(s) in terms of its/their intended purpose(s)."[30]

The Regulation differentiates between different economic operators (Table 6), such as means manufacturers, authorised representatives, importers, and distributors, who are all subject to different obligations.[31]

ECONOMIC OPERATOR	OBLIGATIONS AND RESPONSIBILITIES
Manufacturer	• **Article 10:** Ensure devices are designed, manufactured, documented, and monitored throughout their lifecycle according to the Regulation's safety and performance requirements. • **Article 10a:** Inform authorities and customers of any anticipated supply interruption of a device that could foreseeably result in serious harm to patients or public health.
Authorized Representative	• **Aricle 11:** Act as the legal entity within the Union for a non-EU manufacturer, verifying their compliance, cooperating with authorities, and sharing legal liability for defective devices.
Importer	• **Article 13:** Verify that devices from a third country are compliant before placing them on the market, add their own details to the product, and cooperate in post-market surveillance.
Distributor	• **Article 14:** Act with due care to verify that devices are compliant before making them available, ensure proper storage/transport,

[28] Ibid, Art 2(1)
[29] Ibid, Art 2(2)
[30] Ibid
[31] Ibid, Art 2(35)

ECONOMIC OPERATOR	OBLIGATIONS AND RESPONSIBILITIES
	and cooperate in surveillance and corrective actions.
All Economic Operators	• **Article 15:** Ensure a person responsible for regulatory compliance is available within the organization (for manufacturers) or continuously at their disposal (for authorized representatives and small manufacturers). • **Article 16:** Assume the full obligations of a manufacturer if they place a device on the market under their own name, change its intended purpose, or modify it in a way that could affect compliance. • **Article 25:** Cooperate to ensure traceability of devices throughout the supply chain by being able to identify who supplied them and to whom they have supplied devices.

Table 6: MDR Stakeholder obligations and responsibilities

NIS 2 Directive

The NIS2 Directive (NIS 2) establishes measures intended to achieve a high common level of cybersecurity throughout the Union, to enhance the functioning of the internal market.[32] Whether Partners have to comply with the obligations laid down in this directive depends on whether they fall into the definitions of entities defined in Article 2 and Annex I and II.

Annex I is a reference to Article 3(g) of Directive 2011/24/EU, which encompasses healthcare providers, meaning hospitals that act as partners in LATE-AYA, fall within the scope of the Directive. Research organisations33 may be considered as falling within the scope of the Directive if they are providing services within the European Union and are classified as medium-sized enterprises or larger entities. Alternatively, if their description falls within one of the categories listed in Article 2(2) NIS2, they may also be deemed as falling within the scope of the Directive. Moreover, as stated in paragraph 5(b) of the Article, Member States have the prerogative to implement this directive within the context of educational institutions, particularly in instances where these institutions engage in critical research activities. Consequently, universities that act as beneficiaries in this project and are based in Member States should verify what the national implementation of the NIS2 Directive stipulates. It is incumbent upon other partners to refer to Annex I and II to ascertain their respective scopes.

[32] European Parliament & Council of the European Union. (2022, December 14). Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive). OJL 333, 80–152.

[33] Ibid, Annex II (7.)

NIS2 set the following measures as the minimum level of protection that needs to be implemented by the relevant partners:

- policies on risk analysis and information system security;
- incident handling;
- business continuity, such as backup management and disaster recovery, and crisis management;
- supply chain security, including security-related aspects concerning the relationships between each entity and its direct suppliers or service providers;
- security in network and information systems acquisition, development, and maintenance, including vulnerability handling and disclosure;
- policies and procedures to assess the effectiveness of cybersecurity risk-management measures;
- basic cyber hygiene practices and cybersecurity training;
- policies and procedures regarding the use of cryptography and, where appropriate, encryption;
- human resources security, access control policies, and asset management;
- the use of multi-factor authentication or continuous authentication solutions, secured voice, video, and text communications, and secured emergency communication systems within the entity, where appropriate.[34]

Cyber Resilience Act

The Cyber Resilience Act (CRA) is a regulation laying down rules regarding the cybersecurity of products with digital elements deployed on the internal market.[35] Within the scope of this Regulation, a "product with digital elements" is "a software or hardware product and its remote data processing solutions, including software or hardware components being placed on the market separately."[36] As outlined by the Grant Agreement, not only will research be conducted during the LATE-AYA project, but also specific tangible digital products, which will fall into the scope of this regulation, will be developed. These products include, for example, the intervention planner, multi-functional chatbot, follow-up dashboards, and virtual assistants, or the EHDS-ready SPE.[37]

Partners who are manufacturing, processing, or distributing such products need to align their practices with the obligations laid down in this Regulation (Table 7).

[34] ibid, Art 21(2).

[35] European Parliament & Council of the European Union. (2024, October 23). Article 1. Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act). Official Journal of the European Union, L 2847, 1–52

[36] Ibid, Art 3(1)

[37] this list is non-exhaustive

OBLIGATIONS	MANUFACTURER	IMPORTER	DISTRIBUTIOR
Product design and risk analysis	Yes	No	No
Conformity assessment	Yes	No	No
Technical documentation	Yes	No	No
CE marking	Yes	No	No
Security monitoring and updates	Yes	Yes, if they are aware	Yes, if they are aware
Incident/vulnerability reporting	Yes	Yes, if they are aware	Yes, if they are aware
Market surveillance / Cooperation	Yes	No	No

Table 7: CRA stakeholder obligations

The CRA also lays down articles regulating the conformity of the product wíth digital elements. Aside from the Articles of the Regulation, Partners should take into consideration the following Annexes as they provide important details:

- Annex I: lists all the essential cybersecurity requirements and vulnerability handling requirements;
- Annex II: describes the minimum information and instructions that the product with digital elements should be equipped with for the benefit of users;
- Annex III: differentiates between important products with digital elements;
- Annex IV: differentiates between critical products with digital elements;
- Annex V: provides the information necessary for the EU declaration of conformity;
- Annex VI: lays down a simplified EU declaration of conformity;
- Annex VII: VII lays down all the information that needs to be included in the technical documentation for products with digital elements;
- Annex VIII: describes the conformity assessment procedure.

Digital Services Act (DSA)

The Digital Services Act applies "to intermediary services offered to recipients of the service that have their place of establishment or are located in the Union, irrespective of where the providers of those intermediary services have their place of establishment."[38] The intermediary services can be differentiated into three information society services, namely the (1) mere conduit services, (2) caching services, and (3)

[38] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), Art 2(1), OJ L 277, 27.10.2022, pp. 1–102.

hosting services.[39]

Based on the nature of the digital components developed within LATE-AYA, several elements could potentially fall under the definition of "intermediary service" as outlined, specifically fitting the categories of 'mere conduit' and 'hosting' service (Table 8).

SERVICES	CLASSIFICATION OF DIGITAL ELEMENTS WITHIN LATE-AYA
A Mere Conduit Service (involves the transmission of information in a communication network)	• Edge data collection modules (D3.1) are designed for the continuous collection of survivor data from smartphones and wearable devices. Task T3.2 describes that these modules will "securely send the data to the SPE" (Secure Processing Environment). This dedicated function of transmitting data from the user devices (the edge) to the central cloud repository fits the definition of conveying information through a communication network. • Multi-function conversational interface of the chatbot (D4.1) will be delivered through existing popular platforms, such as WhatsApp, Telegram or Facebook Messenger. While the chatbot provides content, the underlying mechanism involves the transmission of personalized streams of messages, resources, and responses, enacting a dialogue over a conversational interface, which relies on network communication.
A Hosting Service (consists of the storage of information provided by, and at the request of, a recipient of the service)	• The SPE (D3.2) is designed and implemented by CERTH to host the datasets generated during the clinical pilot study. These datasets comprise large amounts of data, including "unconventional patient generated data" from mobile and wearable devices. Since this environment securely stores data (provided by the participants/recipients) at the request of the service it aligns directly with the definition of a hosting service.

Table 8: DSA mapping

Therefore, the obligations imposed on service providers should be taken into account in the design and operation of the digital components described above.

[39] Ibid, Art 3(g).

b. LATE-AYA Ethical Framework

Key Ethical Principles

CHILDREN

As mentioned before, the LATE-AYA project's target group for the clinical study includes adolescents and young adults diagnosed with cancer between the ages of 15 and 39. This age range means that the project will explicitly involve participants who are legally considered minors.

The Grant Agreement acknowledges that the inclusion of minors raises specific and highly sensitive ethical issues that require special attention. Therefore, a robust framework within the project was set up specifically to address these concerns:

1. **Vulnerability and Protection:** A dedicated working group, led by ethics and legal experts, will provide continuous ethical, legal, and social scrutiny of all project activities to offer guidance and ensure the protection of minors.

2. **Informed Consent and Minor's Assent:** The project will implement rigorous protocols for informed consent. While parental consent is required for minors, the project also emphasizes the importance of respecting the minor participant's autonomy and ensuring their own assent is obtained. The process must balance the need for parental involvement with the adolescent's developing decision-making capacity. Both participants and their legal representatives will be fully informed about the study's purpose, risks, benefits, and procedures.

3. **Justification for Inclusion of Minors in Clinical Study:** Minors will only be included in the clinical investigation if there is a scientific expectation that their participation will provide a direct benefit that outweighs any risks and burdens involved, and the research is essential for this age group.

All partners whose work involves interaction with minors must adhere to the applicable national and European laws, which are designed to ensure the utmost protection of underage participants.

ETHICAL CONSENT

Obtaining consent during medical research is one of the key principles of research ethics. The process of informed consent serves two primary purposes: ethical and legal[40].
Firstly, it serves to safeguard patient rights. Secondly, it fosters transparency between healthcare professionals and patients. Thirdly, it serves to promote trust between the

[40] One important element that needs to be pointed out is that ethical consent does not equate consent for the processing of personal data under the GDPR, and as such the project stakeholders must ensure that both perspectives are considered when depending on consent for the project-related activities.

aforementioned parties.[41]

Valid ethical consent must comply with the following criteria:

1. Consent must be obtained before the research begins.
2. Consent must be given voluntarily, with full understanding of the research.
3. Participants must understand risks, benefits, and alternatives.
4. Participants must be aware of the potential consequences of medical interventions.
5. Participants must be empowered to make active decisions about their involvement.[42]

It is important to note that children under the age of 18 are legally unable to provide informed consent for themselves. Consequently, the consent of their legal guardians is required on their behalf. Those responsible for the collection of study participants' data are obliged to consider this aspect and to implement a suitable consent procedure in accordance with these ethical standards.[43]

The principle in question is enshrined in a number of international instruments, including the Declaration of Helsinki, the Universal Declaration on Bioethics and Human Rights, the Oviedo Convention, and the International Covenant on Civil and Political Rights.

MANAGEMENT OF ONGOING CONSENT

The possibility of participants transitioning from minors to adults throughout the course of this project presents a specific challenge for the management of ongoing consent, requiring appropriate procedures for re-consent once the age of majority is reached. A central ethical challenge is to ensure that consent stays valid as participants grow older and their situations, attitudes, and decision-making abilities change. Adolescents, in particular, are still learning to fully understand the consequences of their involvement. This makes it important to handle their consent as it changes over time.

The project establishes the following measures that are responsive to the evolving capacities of the child, ensuring that the child's maturity and ability to provide informed consent are taken into account as they develop.

1. Continuous Right to Withdraw

- All participants are granted the ongoing right to withdraw from the project at any time, without any penalty or negative consequences.

[41] Shah, P., Thornton, I., Kopitnik, N. L., & Hipskind, J. E. (2025, January). Informed Consent. In StatPearls. StatPearls Publishing. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK430827/
[42] Ibid
[43] Ibid

This project has received funding from the European Union's EU4Health Programme under Grant Agreement No 101214326

- This right applies irrespective of the age of the participant and remains valid even when a participant transitions from being a minor to reaching adulthood.

2. Informed Consent as an Ongoing Process

- Informed consent is treated as a continuous dialogue. At study entry, both the child (through age-appropriate assent) and their parent/guardian (through consent) are informed that participation is voluntary and withdrawal is always possible.

3. Ethical Oversight and Monitoring

- A dedicated ethics and compliance working group will continuously monitor project activities to ensure the protection of minors.
- Its oversight responsibilities include verifying that ongoing consent procedures are respected, that evolving capacity is appropriately addressed, and that the right to withdraw is upheld throughout the project lifecycle.

Based on the detailed ethical considerations outlined in the sources for the LATE-AYA project, the project beneficiaries should consider implementing a specific measure to revisit consent at key developmental milestones, such as when a minor participant reaches the age of legal majority.

Overview of Consent and Ethical Approval Status

Table 9 provides a detailed overview of the specific consent requirements, information rights provided to participants, and the status of ethical or IRB approval for each LATE-AYA partner involved in activities requiring ethical consideration at this stage of the project.

PARTNER	PARTICIPANTS' INFORMATION RIGHTS	CONSENT MECHANISM	ETHICAL/REGULATORY APPROVAL
INT	Participants receive verbal information and a written subject sheet in the local language, prepared per ICH-GCP. The informed consent form meets applicable data-protection requirements.	Written, dated consent from both patient and investigator is required before any study procedures.	Mandatory. The trial may begin only after the site's independent ethics committee has reviewed and approved all required documentation.
IEO	Participants are informed about data processing, including statistical analyses and AI model development, through a standardized, locally translated informed consent form and verbally.	Signed consent is obtained, dated and documented at each site before any data collection or processing	The trial will be initiated only after all required legal documentation has been reviewed and approved by the responsible independent ethical committee (EC) of the center according to all applying national and international regulations.
VHIO and VHIR	Medical visits or telephone calls as well as the information consent form for patients.	Specific HIP-CI from the hospital	Mandatory, before data collection
NVI	Participants are informed under voluntary basis	Paper based consent form will be filled and signed by the participant. These will be	The IRB approval will be obtained.

PARTNER	PARTICIPANTS' INFORMATION RIGHTS	CONSENT MECHANISM	ETHICAL/REGULATORY APPROVAL
		stored at NVI's clinical trial unit	
DOT	Participants are informed through site-specific consent forms approved by Ethics Committees and are told they can withdraw consent at any time.	Can withdraw the consent for their mobile and device data being used for research purposes at any time.	Processing activities are covered by the ethics approvals obtained by recruiting clinical sites; DOTSOFT operates as a data processor under these approvals
VS	Consent to data processing will have been already granted by patients by the time these activities take place	Not provided (relying on clinical site consent)	Not needed.
UTW	Participants were informed about the project purpose, voluntary participation, and data use in the survey introduction and workshop invitations.	Electronic consent (checkbox via Qualtrics) and written consent for workshop sessions.	Ethical approval obtained from the UTEW BMS Ethics Committee (Ref. No. 251877, July 2025) covering survey and workshop data. No additional approval is required unless new data types (e.g., clinical or patient-level data) are collected in the future.
SAM	Participants were informed about the project purpose, voluntary participation, and they signed a consent form in advance.	Electronic consent collected (checkbox before participation) by the Data-Collection App.	Not applicable
CQ	Participants were informed about the project purpose.	Electronic consent collected.	Not applicable
UNITO	A subject information sheet in the local language and prepared in accordance with the ICH Note for Guidance on Good Clinical Practice (ICH, Topic E6, 1995) will be provided for the purpose of obtaining informed consent. In addition to this written information, the investigator or the designate will inform the patient verbally.	The consent of the subject must in writing before activities are carried out. The consent must be signed and personally dated by the patient and by the investigator or a designated person. Provision of consent will be confirmed in the eCRF by the investigator	The trial will be initiated only after all required legal documentation has been reviewed and approved by the responsible independent ethical committee (EC) of the center according to all applying national and international regulations.
KTU	Participants were informed through participant information sheets, privacy notices, and/or in-app information screens describing the purpose of processing, categories of data collected, legal basis, retention, data-sharing conditions, and their GDPR rights	Consent is or will be collected in written or secure electronic form	Ethical approval will be obtained from the relevant institutional ethics committee before any data collection. All procedures for informed consent, data processing, and handling of sensitive health information follow GDPR, national regulations, and the approved ethical protocol.
CERTH	N/A, CERTH does not interact with participants. Only processes data collected under the Ethics approvals and informed consent obtained by the recruiting clinical sites.	N/A, CERTH does not interact with participants.	CERTH notes that it will investigate the need for specific ethical approval later, once more details about the processing activities are available.

Table 9: Consent requirements and approach

Instruments Encompassing Ethical Principles

The following instruments should be taken into consideration as they provide ethical principles relevant to the LATE-AYA project.

1. GENERAL ETHICAL PRINCIPLES

General principles set the basic rules for protecting human dignity and rights in research. They promote equality, non-discrimination, privacy, autonomy, and the right to health and well-being. They also highlight justice, fairness, solidarity, and the need to protect vulnerable groups. These principles act as essential safeguards in all scientific and healthcare settings.

Relevant instruments:

- Universal Declaration of Human Rights[44]
- European Convention on Human Rights[45]
- Charter of Fundamental Rights of the EU[46]
- International Covenant on Civil and Political Rights Ethical Principles[47]
- International Covenant on Economic, Social, and Cultural Rights: Ethical Principles[48]

2. MEDICAL AND CLINICAL PRINCIPLES

Medical principles focus on protecting participants in health-related research. They ensure respect for individuals, informed consent, accountability, and the protection of dignity, autonomy, and privacy. These guidelines help researchers balance risks and benefits, maintain transparency, promote fairness in including participants, ensure independent ethical review, and make sure that medical studies are scientifically valid and ethically justified.

Relevant instruments:

- WMA Declaration of Taipei on Ethical Considerations regarding Health Databases and Biobanks Ethical Principles[49]

[44] European Union. (2000, December 18). Charter of Fundamental Rights of the European Union (2000/C 364/01). Official Journal of the European Communities. Retrieved from https://www.europarl.europa.eu/charter/pdf/text_en.pdf

[45] Council of Europe. (1950, November 4). European Convention on Human Rights, as amended by Protocols Nos. 3, 5, 8, 11, 14, and 15 (ETS No. 5). Entered into force September 3, 1953. Council of Europe. Retrieved from https://www.echr.coe.int/documents/d/echr/convention_ENG

[46] United Nations General Assembly. (1948, December 10). Universal Declaration of Human Rights (GA Res. 217A (III)). Adopted by the General Assembly on 10 December 1948. United Nations. Retrieved from https://www.un.org/en/about-us/universal-declaration-of-human-rights

[47] United Nations General Assembly. (1966, December 16). International Covenant on Civil and Political Rights (GA Res. 2200A (XXI), UNTS Vol. 999, p. 171). Entered into force March 23, 1976. United Nations. Retrieved from https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

[48] United Nations General Assembly. (1966, December 16). International Covenant on Economic, Social and Cultural Rights (GA Res. 2200A (XXI), UNTS No. 993, p. 3). Entered into force January 3, 1976. United Nations. Retrieved from https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights

[49] World Medical Association. (2016, October). Declaration of Taipei on Ethical Considerations regarding Health Databases

- Declaration of Helsinki[50]
- International Ethical Guidelines for Health-related Research Involving Humans[51]
- Universal Declaration on Bioethics and Human Rights Ethical Principles[52]
- Oviedo Convention[53]

3. ARTIFICIAL INTELLIGENCE (AI) PRINCIPLES

AI frameworks set standards for the responsible design and use of intelligent systems. They ensure that AI remains lawful, ethical, and technically robust. These principles also promote human oversight, fairness, transparency, and accountability throughout the AI lifecycle.

Relevant instruments:

- Ethics Guidelines for Trustworthy AI[54]
- Fair AI Attribution Framework[55]

4. DATA PROTECTION AND PRIVACY PRINCIPLES

Data protection frameworks offer rules for the legal, fair, and clear handling of personal data. They guarantee data quality, security, and accountability. These frameworks also encourage ethical reuse and responsible data management across different regions and industries.

Relevant instruments:

- UN Personal Data Protection and Privacy Principles[56]
- OECD Good Practice Principles for Data Ethics in the Public Sector[57]

---

and Biobanks (Revised version). Adopted by the 67th WMA General Assembly. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/

[50] World Medical Association. (2024, October). Declaration of Helsinki: Ethical principles for medical research involving human subjects. Adopted by the 75th World Medical Association General Assembly. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki/ en.wikipedia.org+15

[51] Council for International Organizations of Medical Sciences. (2016). International Ethical Guidelines for Health-Related Research Involving Humans (4th ed.) [Prepared in collaboration with the World Health Organization]. Geneva: CIOMS. Retrieved from https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf

[52] United Nations Educational, Scientific and Cultural Organization. (2005, October 19). Universal Declaration on Bioethics and Human Rights. Adopted by the UNESCO General Conference. Retrieved from https://www.unesco.org/en/legal-affairs/universal-declaration-bioethics-and-human-rights

[53] Council of Europe. (1997, April 4). Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (European Treaty Series No. 164). Council of Europe. Retrieved from Council of Europe Treaty Collection https://rm.coe.int/168007cf98

54 High-Level Expert Group on Artificial Intelligence. (2019, April 8). Page 2. Ethics Guidelines for Trustworthy AI. Publications Office of the European Union. Retrieved from https://ec.europa.eu/digital-strategy/en/library/ethics-guidelines-trustworthy-ai

[55] KIA Digitalisering, Digital Holland. FAIR AI Attribution (FAIA): Setting clear standards for human-created and AI content. https://kia-digitalisering.com/news/fair-ai-attribution-faia-setting-clear-standards-for-human-created-and-ai-content#:~:text=FAIA%20is%20developing%20an%20open%20framework%20for%20AI%20attribution%20to,creation%20is%20attributed%20and%20justified.

[56] United Nations High-Level Committee on Management. (2018, October 11). Principles on Personal Data Protection and Privacy for the United Nations System Organizations. Adopted by the High-Level Committee on Management. United Nations. Retrieved from https://www.unsystem.org/privacy-principles

[57] Council of Europe. (1981). Convention for the Protection of Individuals with regard to Automatic Processing of Personal

5. PRINCIPLES FOR THE PROTECTION OF MINORS

Given the vulnerability of minors, certain international conventions emphasize the importance of putting the child's best interests first in all activities. Children's rights to health, protection, and specific safeguards are key. This ensures that their involvement in research honors their unique status and needs.

Relevant instruments:

- UN Convention on the Rights of the Child[58]

Data (ETS No. 108).

[58] United Nations General Assembly. (1989, November 20). Convention on the Rights of the Child (General Assembly Resolution 44/25). United Nations. Entry into force September 2, 1990. Retrieved from https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child ohchr.org

7. Other Issues

In addition to the basic obligations deriving from the Horizon Europe Grant Agreement (GA), the LATE-AYA project will implement the following European, international, and sector-specific procedures concerning the management of data and research outputs:

- All statistical analyses for the clinical trial will be conducted in accordance with internationally recognized guidance documents, specifically the ICH E9 Statistical Principles for Clinical Trials.
- All activities will comply with the national legal and ethical requirements of the country where they take place. For example, the clinical centres drafting the study participant consent form must do so according not only to the EU law and standards but also according to local/national requirements.
- The research design accounts for the upcoming need for AI device certification by adopting standards from the ISO/IEC JTC 1/SC42 committee (for Artificial Intelligence, Risk Management, and Trustworthiness).
- The Europrivacy GDPR compliance assessment methodology will be incorporated into the LATE-AYA project to ensure a standardized and methodologically sound approach to compliance across partners.
- Clinical activities, including the study intervention, will be conducted in accordance with NCCN guidelines (v3.2021)

8. Conclusion and Future Work

This document represents the first official version of the LATE-AYA Data Management Plan (DMP), as described in deliverable D1.4 (Data management plan v1). It sets the basic framework for managing all research data produced by the project, especially focusing on the sensitive health and psychosocial data collected from adolescent and young adult (AYA) cancer survivors. The information provided in this document is based on the answers from the Data Management Questionnaires provided by the project beneficiaries, together with the information from the Grant Agreement.

The framework's two primary pillars are strict adherence to all applicable EU laws, including the GDPR, EHDS, and AI Act, as well as any relevant ethical standards. All measures outlined are tied to the project's technical setup, particularly the Secure Processing Environment, which integrates privacy by design and by default. In parallel, the project partners should be developing strategies for standardizing metadata and improving data quality to ensure that all produced datasets follow the FAIR principles. This should maximize their value for future use while minimizing risks to data subjects. Looking forward, this DMP is meant to be a living document.

As part of the "Quality of life (AYA)" cluster, the project must maintain alignment with the common data management goals established through the Common Chapter for Cluster Data Management Working Group, with the final goal of enabling the integration of project datasets onto the UNCAN.eu platform and to prepare the multisource

datasets for future cataloguing with relevant Health Data Access Bodies (HDABS). In this context, it is recommended that all project partners consider the need to translate the principles defined in this document onto concrete, operational foundational deliverables that will enable project activities. This includes (but is not limited to) the contractual framework alignment, ethical oversight and compliance actions, and the refinement and implementation of the Secure Processing Environment by relevant partners.

Future versions of this deliverable are planned culminating in D1.5, Data management plan v2, which is due in month 58. These will provide more details on operational practices and technical implementations as the project develops. As such, this document will reflect the project's compliance measures and its contribution to strong, secure, and reusable data practices for European survivorship research, especially through its planned integration with the UNCAN.eu platform.

Annex 1: Cluster Common Chapter on DMP

Common Data Management Plan Chapter

Document date: 26-11-2025
Document version: V1

"Quality of Life (AYA)" Cluster Projects funded under HORIZON-MISS-2024-CANCER-01-05 & HORIZON-MISS-2023-CANCER-01-04		
Project Title	**Acronym**	**Grant No.**
e-health tools to promote Equality in Quality of Life for childhood to young adulthood cancer patients, survivors and their families - a Pan-European project supported by PanCare and Harmonic consortia	e-QuoL	101136549
Understanding and addressing LATE-effects of treatment of AYA cancer survivors with AI-based digital phenotyping and non-invasive holistic approach	LATE-AYA	101214326
smart Mirrors supporting healthier lives of Adolescents and Young Adults after cancer	MAYA	101213323
Patient-centred, evidence-based late effect screening and follow-up care of survivors of adolescent and young adult cancer	PanCare4AYA	101213107
Prediction and prevention of late effects in AYA cancer survivors – An effort to understand, predict and prevent late effects in AYAs 15-39 years of age, with a focus on fertility and gonadal toxicity	PredictAYA	101214879
Co-creation and implementation of an intervention to improve the understanding of TesticulaR cANcer late effects and unmet Supportive CarENeeDs of AYA survivors using eXtended Reality	TRANSCEND-XR	101214413

Funded by the European Union

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 European Health and Digital Executive Agency (HADEA). Neither the European Union nor the granting authority can be held responsible for them. AYA QoL as a cluster includes the following projects: e-QuoL, LATE-AYA, MAYA, PanCare4AYA, PredictAYA, TRANSCEND-XR.

1. Quality of Life (AYA) Cluster Data Management

1.1. Introduction and Cluster Vision

The **"Quality of Life (AYA)" Cluster** brings together six EU-funded projects dedicated to advancing the understanding and management of late effects in adolescents and young adults (AYA) with cancer. This includes the five projects funded under the call *HORIZON-MISS-2024-CANCER-01-05 – Improving the understanding and management of late-effects in adolescents and young adults (AYA) with cancer*, as well as the **e-QuoL** project (funded under *HORIZON-MISS-2023-CANCER-01-04)* which officially joined the Cluster in September 2025.

Through coordinated clustering activities, the initiative will make a substantial contribution to strengthening AYA cancer care across Europe. By fostering collaboration, avoiding duplication of efforts, and promoting synergies, the Cluster aims to enhance policy impact and engagement from key stakeholders, experts, and policymakers, ultimately driving more cohesive and effective outcomes for AYA cancer survivors.

Projects in the Cluster:

- **e-QuoL**, Grant Agreement ID: 101136549 (01 January 2024 – 31 December 2027)
- **LATE-AYA**, Grant Agreement ID: 101214326 (01 June 2025 – 31 May 2030)
- **MAYA**, Grant Agreement ID: 101213323 (01 June 2025 – 31 May 2029)
- **PanCare4AYA**, Grant Agreement ID: 101213107 (01 April 2025 – 31 March 2030)
- **PredictAYA**, Grant Agreement ID: 101214879 (01 June 2025 – 31 May 2030)
- **TRANSCEND-XR**, Grant Agreement ID: 101214413 (01 June 2025 – 30 November 2030)

The **main objective** of this chapter is to establish a common Data Management Plan (DMP) framework within the "Quality of life (AYA)" cluster. This framework aims to ensure data interoperability, validation, and protection, as well as foster data exchange among projects, with a particular focus on preparing for the future UNCAN.eu and other federated data platforms. This preparation specifically involves addressing commonalities in data standards, data validation, data protection, and facilitating data exchange.

However, the UNCAN.eu platform will be implemented through the **UNCAN-CONNECT project**, which started in September 2025. Because the initial integration requirements for this platform are not yet defined and are not expected to be available for at least three years, there are currently no concrete requirements for this integration. UNCAN blueprint is already available and this working group will take the suggestions there into consideration.

Given this timeline, the methodology for integrating the projects' data approaches with UNCAN-Connect will be dynamic and evolving. The Working Group will adopt an iterative strategy, monitoring the outputs of UNCAN-Connect as they become available to progressively align the cluster's standards with the emerging requirements. This flexible approach ensures that, while starting with the Blueprint recommendations, the integration pathway will adapt over time to match the final technical specifications of the federated platform.

For at least two of the cluster projects (PredictAYA and PanCare4AYA) the development of DMPs envisions also interactions with consortia resulting from these projects:

UNCAN-CONNECT (ID: 101215206) - UNCAN-Connect project will establish a decentralised collaborative network for cancer research, focusing on paediatric cancers, lymphoid malignancies, pancreatic, ovarian, lung, and prostate cancers. The initiative will create a governance framework and engage various stakeholders, including researchers, SMEs, and citizens, while facilitating data storage, access, and sharing across member states.

CANDLE - National CAncer data Node DeveLopErs (ID: 101214368) - CANDLE project will enhance national and international health data infrastructures, aligning with the European Health Data Space (EHDS) in Member States. It will also identify and address barriers to implementing the UNCAN.eu and ECPDC digital platforms.

EU-CIP - EUropean Cancer Information Portal (ID: 101214125) - EU-CIP project will create the European Cancer Patient Digital Centre (ECPDC) Information Portal to enhance health literacy and improve access to cancer care information across Europe.

1.2. Cluster Data Management – Plan

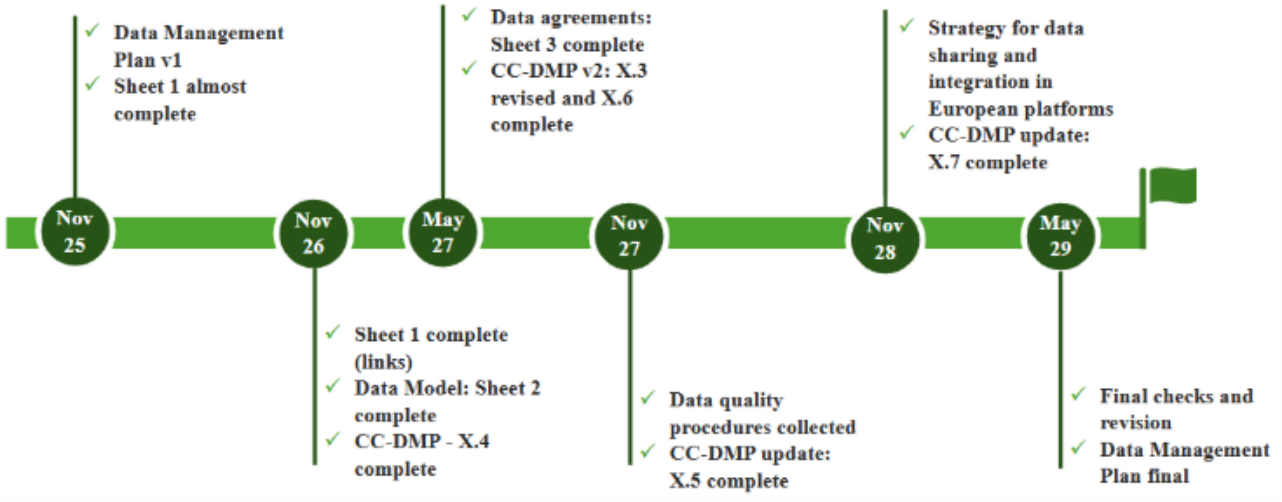
The cluster has been organized in a step-by-step approach to ensure a smooth and effective process:

1. With the leadership of the European Cancer Organisation (ECO), a Coordination Body for cluster data management has been established: **"Common Chapter for Cluster Data Management Working Group" (CCDMP-WG)**. It has been suggested by the leading partner of the CCDMP-WG that members should include the technical manager from each participating project, data protection officers (DPOs) and clinical leads. In an internal shared list, one person (or group of persons) has been appointed as the lead coordinator for each participating Cluster project and will ensure that corresponding members within their project are working in this activity.
2. ***A* Kick-off meeting was** held on Thursday 30th October at 11CEST to agree on the work plan, timeline, and the structure of the chapter and shared spreadsheets. A follow-up meeting was held (Thursday 13th November at 10CEST) also to make final agreements after leaving time to all projects to review the draft common chapter and provide the necessary information to fit the overall timeline for the WG.
3. The WG will work gathering some particular and agreed information, organized in templates "**Shared excel sheets for information collection.**"

All projects participating in the CCDMP-WG have reviewed and agree to provide the information indicated in the shared excel sheets to collect the necessary information in a standardised format.

4. **Analysis and drafting.** The CCDMP-WG will collect the information in the shared sheets (sheet 1 and mainly sheets 2 and 3) and analyses the collected information to identify commonalities, differences, and challenges.

In the final release of this deliverable, the project leading the CCDMP-WG will draft the common chapter, with iterative input and review from the working group.
The following timeline will be followed to feed and review the information (as per the shared sheets) by all projects. Deadlines for the deliverables related to the information gathered have been taken into account for each particular project.



1.3. Project Data Governance

e-QuoL: A Data Management Plan (DMP) has been established (Month 6) and will be updated by Month 48 to ensure the project complies with **FAIR principles** (Findable, Accessible, Interoperable, Reusable) and **Horizon Europe** standards. The plan covers the entire lifecycle of qualitative, quantitative, and health-related data, detailing protocols for curation, storage, IP protection, and dissemination via public repositories (e.g., Zenodo). Strict **GDPR compliance** is enforced through local anonymization and rigorous ethical adherence, with **HUGO** leading project-level management and Work Packages (WPs) handling local data quality.
Timeline & Objectives: developed at M6 with an update scheduled for M48; aims to guarantee data quality, accessibility, and reuse.

- Scope & FAIR Compliance:
  - Data Types: Survey, focus group, and health-related data.
  - FAIR: Use persistent identifiers, open access post-publication (using Creative Commons licenses), and standardized metadata.
  - Storage: Defined methods for long-term archiving and backup.
- Legal & Ethical Compliance:
  - GDPR: Data is anonymized locally; keys remain at source.

  - Standards: Aligned with EU directives (ePrivacy, GDPR) and Horizon Europe guidelines.
  - Management: HUGO leads project oversight; local WPs manage specific analysis and storage costs.

- Alignment Recommendations:
  - Explicitly detail anonymization processes and key management.
  - Ensure data sharing agreements cover restricted datasets.
  - Clarify licensing terms for all data types.

Late-AYA: The LATE-AYA project has adopted a comprehensive and integrated approach to data governance, ensuring strict compliance with European regulations while making high-quality, sensitive data available for reuse. This strategy is implementation-oriented through mandatory technical architectures, dedicated roles, and specific deliverables integrated into the overall project workflow, which are being further refined in the project's DMP.

I. Foundational Legal Framework & Compliance (Mandatory Actions)

- Regulatory Adherence: Adhere strictly to the General Data Protection Regulation (GDPR) and applicable national data protection laws, particularly by aligning Data Protection Practices with the EU GDPR Certification (Europrivacy) checklists.
- Future-Proofing Data Sharing: Align the project's data environment and governance framework with the forthcoming requirements of the European Health Data Space (EHDS) Regulation.
- Formalizing Agreements: Cooperate to fulfill legal obligations and formally establish all data usage protocols by concluding separate mandatory agreements, such as data processing, data sharing, and/or joint controller agreements. This includes drafting Data agreements templates (D2.3, led by UDGA) and obtaining Data agreements signed (D6.3, led by INT) before starting the clinical studies.
- Ethics-by-Design: Elicit legal and ethical requirements early in the Co-creation (WP2) phase (T2.3), ensuring the LATE-AYA framework is designed with legal and ethical constraints in mind, especially concerning minors and the use of AI systems.

II. Technical Implementation & Data Lifecycle

- Secure Processing Environment (SPE) Requirement: Design and implement a cloud-based Secure Processing Environment (SPE) (D3.2, led by CERTH) to host collected datasets.
- Architecture: The SPE must be designed according to Zero Trust Architecture (ZTA) principles to ensure maximum protection of personal data.
- Data Protection Mechanisms: Utilize methods like encryption, access controls, anonymization, and pseudonymization to prioritize privacy-by-design and security-by-design. Align with best practices, administrative guidance by national Data Protection Authorities, and relevant standards to this end.

- Data Planning and Quality: Establish and regularly update the Data Management Plan (DMP) (D1.4, D1.5), led by the Associated Partner UDGA, ensuring adherence to FAIR principles (Findability, Accessibility, Interoperability, and Reusability).
- Data Reuse: Establish a process to share high-quality datasets (D7.1) and AI models (D7.2) by ensuring their readiness for effective integration into the UNCAN.eu platform (T10.5) and for cataloging with future Health Data Access Bodies (HDABs).
- Data(bases) IPR protection & exploitation analysis: Collaboration with relevant enablers, such as the EU Exploitation Booster Service and the EU IPR Helpdesk to ensure proper protection and maximize sustainable exploitation of project outputs, including datasets (whenever relevant and viable given the nature of the data) is expected.

III. Integrated Governance and Oversight

- Project Management Integration: The Project Coordination framework established in the Grant Agreement includes key roles dedicated to governance aspects in WP1:
  - A Quality & Risk Manager overseeing data management and producing the DMP.
  - An Ethics Manager responsible for ethical and privacy concerns.
- Steering Board Composition: The highest decision-making body, the Steering Board, is mandated to include Ethics and legal experts (from UGR and UDGA) to monitor project progress and compliance alongside scientific and technical managers, as well as to consider the views of an ethics advisory board to guarantee continuous ethical, legal and social scrutiny of activities throughout the project's execution.
- Contractual Flow-down: Ensure that essential contractual obligations (including proper implementation, conflict of interests, confidentiality, security, ethics, and record-keeping) are applied not only to Beneficiaries but also to Associated Partners (UDGA).

Specific & operationalization-oriented information is currently being developed by the consortium and will be specified in the DMP and follow-up activities. Alignment with UNCAN and other best practices is expected.

MAYA: The MAYA project will generate and, in some cases, re use multimodal datasets collected from diverse sources during Living Labs and pilot site activities. These include data from IoT health devices, smart mirrors, wearable sensors, conversational AI tools, medical imaging, and patient reported outcomes. The data types cover both clinical and behavioural health information (e.g., cardiovascular metrics, treatment history, lifestyle tracking), system telemetry for engagement/adherence monitoring, and managerial/ethical documentation.
Data will be produced in human and machine readable formats (e.g., RDF, XML, JSON, CSV) and will adopt healthcare interoperability standards such as HL7 FHIR for clinical data. All datasets will be managed under FAIR principles and GDPR compliance, with pseudonymization / anonymisation applied as appropriate. The collected information will support the project's objectives of characterising late cardiovascular effects in adolescent and young adult cancer survivors, evaluating the MAYA iCARE health hub, and deriving participatory, evidence

based prevention and follow up recommendations.

PanCare4AYA: As PanCare4AYA coordinator, the Princess Máxima Center for Pediatric Oncology (PMC) will maintain oversight of all project activities, including data collection, documentation and analysis, long-term preservation and archiving of project data, as well as management of legal agreements with data users to ensure regulatory compliance. The Netherlands Cancer Institute (NKI) oversees participant recruitment and data collection in the PanCare4AYA prospective cohort study, which will test the PanCare4AYA intervention – the *AYA Cancer Survivor Screen* programme – in 20 clinical sites in 11 countries. Before study launch, each clinic needs to conduct a data protection impact assessment (DPIA) and obtain ethical approval. In addition to the DMP, a data protection plan will be developed and implemented to safeguard compliance with national and EU data protection laws and regulations (e.g., GDPR). Study data will be collected locally from study staff at each clinical site and centrally managed by the NKI. The NKI will set up and maintain a common database using the GDPR-compliant platform Castor EDC, through which study questionnaires will be distributed to participating AYA cancer survivors. A data dictionary will detail common terminologies and coding rules to ensure data is collected and prepared according to a standardised format. In addition, PanCare4AYA will set up a Study Management Group with local centres and the NKI to monitor data collection and conduct regular and systematic quality checks for completeness, plausibility and consistency.

PredictAYA: Karolinska Institutet (KI), coordinator of PredictAYA project, maintains overall oversight of data governance, harmonisation, documentation, and regulatory compliance across the consortium. The coordination of the DMP and its integration with ethical and legal requirements under Work Package 9 (Ethics) is led by Civitta Foundation (CIVITTA), while Work Package 8 (Clinical Studies Coordination) oversees the operational management, including ethical permits and progress of all clinical studies within the project.
PredictAYA generates multi-layered research data from five complementary studies, each addressing distinct but interlinked aspects of long-term outcomes in AYA cancer survivors:
A – an epidemiological follow-up study using population-based registry data to examine the incidence, survival, and late effects of cancer in AYA survivors across Europe.
B – a retrospective data-analysis study integrating genetic and other biobank data to model risk factors and molecular mechanisms underlying treatment-related late toxicities.
C – a prospective multi-centre observational cohort study investigating ovarian toxicity and fertility outcomes following cancer therapies in young female survivors.
D – a prospective multi-centre observational cohort study assessing the frequency and prediction of chemotherapy-induced testicular dysfunction in male AYA survivors.
E – a retrospective data-assembly study focused on the prediction of chemotherapy-induced ovarian and testicular dysfunction, including patient-reported outcomes.
Prospective clinical and patient-reported outcome data are collected locally and assembled through REDCap, a electronic data-capture system hosted at KI. Retrospective and registry data are re-used under approved national frameworks and research derogations. Each partner institution acts as a data controller for datasets collected or held under its responsibility, ensuring pseudonymisation at source, secure local storage, and access restricted to authorised personnel.

TRANSCEND-XR: The TRANSCEND-XR project operates under a rigorous data governance

framework overseen by University College Cork, which ensures regulatory compliance across the consortium through three distinct Data Management Plans scheduled between November 2025 and November 2030. Ethical oversight is led by Universidade Católica Portuguesa to ensure that all collected data—ranging from surveys and World Café workshops to immersive XR interactions—are pseudonymised or anonymised prior to processing. To maintain high security standards, qualitative data used for co-creation is stored in secure institutional repositories, and all data transfers between partners are conducted through encrypted, access-controlled environments that fully comply with EU data security standards.

Strategically, the cluster aims to achieve interoperability with the UNCAN-CONNECT blueprint by developing aligned Data Sharing Agreements and consent templates that adhere to Article 9(2)(j) of the GDPR for scientific research. A cross-project Legal and Ethical Working Group actively monitors regulatory updates to promote procedure convergence, ensuring that the federated architecture meets the highest legal standards. Particular attention is dedicated to the ethical management of behavioural data collected in XR environments, with strict protocols in place to safeguard participants' autonomy, dignity, and emotional safety throughout the project.

1.4. Common Data Landscape

In this chapter, each data model responsible for each project will work to summarize the data that will be used in the project. Initial work will be to complete the tables based on the WG-timeline, and then to support on the final analysis. The common data landscape will consider:

- **Data types and sources:** main types of data being generated across the cluster (e.g., clinical data, patient-reported outcomes, genomic data, imaging data).
- **Data standards and interoperability:** This is a key section, to:
  - Identify the common data models (e.g., OMOP-CDM, i2b2) and terminologies (e.g., SNOMED CT, LOINC, ICD-10) used by projects.
  - Describe the strategy for harmonization. This could involve mapping project-specific data to a common standard or defining a minimum set of core data elements for the cluster.
- **Metadata standards:** metadata standards that will be used to describe the datasets, ensuring they are **Findable** and **Accessible** (the 'F' and 'A' in FAIR; 'I' and 'R' are covered with the standards and interoperability approaches).

Details and specific tracking of these datasets descriptions will be in the associated ***'Data asset inventory' sheet***.

1.5. Data Quality and Validation

To ensure the reliability and interoperability of all information managed by the cluster, all participating projects will follow their own quality rules but with adherence to a common framework for data quality control and validation. The assessment of data quality will rely on a structured evaluation of standardized dimensions[1], such as completeness, consistency, validity, accuracy, and data provenance.

Within this framework, particular emphasis will be placed on the principles of missingness, plausibility, and completeness, which are fundamental to ensure that datasets are technically sound and analytically useful.

- Missingness: evaluates the presence and relevance of missing data, ensuring that any absence of information is documented, explained, and manageable for the intended use.
- Plausibility: verifies that recorded data are realistic, internally coherent, and compatible with expected patterns or known reference values.
- Completeness: ensures that all necessary information required for analysis or interoperability is captured and adequately described.

Each project will undergo their own standardized quality checks using measurable indicators aligned with these dimensions and expert-agreed thresholds.
A data consistency check protocol is encouraged also to be systematically applied. This protocol might include:

1. Cross-record verification to identify contradictory or duplicated information across datasets.
2. Format and standard compliance checks to ensure that data adhere to agreed schemas, terminologies, and technical specifications.
3. Validation against reference sources where applicable, to confirm the correctness and temporal coherence of key variables.
4. Automated and manual plausibility controls to detect unrealistic or outlier values that may indicate data entry or transformation errors.

By passing these quality and consistency validations to the project datasets will ensure the readiness of each for integration and exchange. This approach guarantees that all shared data are accurate, complete, coherent, and fit for their intended analytical or operational purpose, ensuring interoperability and trust across the cluster and with external platforms.
The data model responsible of each project will be committed to these activities.
Details and specific tracking of these datasets descriptions will be in the associated ***'Data asset inventory' sheet***.

1.6. Legal and Ethical Framework

This chapter summarizes the comprehensive legal and ethical framework established to govern data usage and compliance across the projects, and will seek to ensure forward-looking alignment on sensitive issues, including ethical approvals, risk categorization and data protection approaches, building upon emerging best practices and enablers such as the European Data Protection Seals.An umbrella framework will seek to outline the specific measures adopted by each project partner to ensure full alignment with General Data Protection Regulation (GDPR) standards.

Key components of this framework will include:

- **Data protection strategy**: How each project addresses strictly mandated data protection requirements under GDPR.
- **Secondary use considerations**: How each project addresses reuse of personal medical data for research, including national law and GDPR interactions with relevant EDPB guidance.
- **Patient consent**: To support future data exchange, the Cluster will build upon the project's practices and the recommendations of relevant national authorities to enable a convergent understanding and a shared overview of consent formulations and opportunities for future alignment.
- **Privacy preservation**: The established strategies for data pseudonymisation or anonymisation to protect data subject identities during processing and exchange will be explored and aligned with emerging jurisprudential developments, standards and solutions.
- **Governance documentation**: Details regarding the Data Sharing Agreements (DSAs), Joint data controllership agreements, and other necessary contractual actions implemented to facilitate authorized data exchange will be addressed jointly by the Cluster partners, to seek the simplification of the development, negotiation and signature process

Granular details and specific tracking of these compliance measures will be found in the associated ***'Legal and ethical framework' sheet***.

1.7. Strategy for data sharing and integration in European platforms

A strategic approach for sharing project data and integrating it effectively with wider European initiatives will be covered in this chapter. This strategy aims to outline the selected technical mechanisms for data exchange, addressing whether a federated query system (where data remains local) or deposition into a centralized repository (such as EHDS) will be utilized.

Main outcomes expected in this chapter as the final release are:

- **Integration strategy:** The specific methodologies for aligning and integrating project outputs with major European platforms, including UNCAN.eu, CANDLE, and EU-CIP.
- **Implementation roadmap:** A roadmap for achieving the necessary technical and legal readiness to facilitate this data exchange and full integration with the targeted external infrastructures.

late aya

Sheet 1: Project & Data Overview

This sheet captures high-level information to understand the scope of each project. Summary here. Links will be added when available.

Project Acronym	Project Title	General Description of Data Generated	Link to Project-Specific DMP	Link to Project-Specific Data model
e-QuoL	e-health tools to promote Equality in Quality of Life for childhood to young adulthood cancer patients, survivors and their families - a PanEuropean project supported by PanCare and Harmonic consortia	The e-QuoL project generates a diverse range of qualitative and quantitative data through surveys, co-creation workshops, and a prospective cohort study to enhance supportive care for childhood and young adult cancer survivors. Work Packages 2 and 3 focus on foundational data, assessing user needs and barriers to develop specifications for the electronic Patient Self-Care Tool (e-PSCT) and adapted digital passports. This provides the basis for Work Package 4, which produces data on technical integration, interoperability standards (such as FHIR and SNOMED CT), and the scalability of these deployments across clinical centers. Finally, Work Package 5 validates the intervention through a multi-country cohort study involving 240 participants, generating critical clinical effectiveness data, patient activation measures, and implementation of metrics based on the RE-AIM framework.	[Link]	[Link]
LATE-AYA	Understanding and addressing LATE-effects of treatment of AYA cancer survivors with AI-based digital phenotyping and non-invasive holistic approach	Data comming from Late-AYA studies, in particular from AYA cancer survivors. Data collected include: clinical information, demographics, baseline clinical evaluation, cancer family history, histology, cancer therapy history, medication and medical events review, physical/emotional/social status, and health-related QoL and LE-specific questionnaires.	[Link]	[Link]
MAYA	Smart Mirrors supporting healthier lives of Adolescents and Young Adults after cancer	The MAYA project will generate and, in some cases, re use multimodal datasets collected from diverse sources during Living Labs and pilot site activities. These include data from IoT health devices, smart mirrors, wearable sensors, conversational AI tools, medical imaging, and patient reported outcomes.	[Link]	[Link]
PanCare4AYA	Patient-centred, evidence-	PanCare4AYA will generate both qualitative and	[Link]	[Link]

Project Acronym	Project Title	General Description of Data Generated	Link to Project-Specific DMP	Link to Project-Specific Data model
	based late effect screening and follow-up care of survivors of adolescent and young adult cancer	quantitative data in the context of the PanCare4AYA intervention, the AYA Cancer Survivor Screen programme, which will be tested in 20 clinical centres in 11 countries to improve AYA Cancer Survivorship Care. During the multi-country implementation study (a prospective cohort study), quantitative and qualitative data from AYA cancer survivors (target: 1,000 participants) and HCPs will be collected at four time points over an 18-month period. This includes several patient-reported outcome measures (PROs) and patient-reported experience measures (PREMs) via online questionnaires, data on clinical outcomes, delivered care, screening and follow-up, the feasibility of programme implementation, participants' experiences with the use of digital tools and healthcare costs related to the programme.		
PredictAYA	Prediction and prevention of late effects in AYA cancer survivors – An effort to understand, predict and prevent late effects in AYAs 15-39 years of age, with a focus on fertility and gonadal toxicity	The PredictAYA project generates multi-level research data combining clinical, registry, biobank, genomic, and patient-reported information.	[Link]	[Link]
TRANSCEND-XR	Co-creation of an Intervention to improve the understanding of TesticulaR cANcer late effects and unmet Supportive CarE NeeDs of Adult and Young Adult survivors using eXtended Reality (TRANSCEND-XR)	TRANSCEND-XR generates a combination of qualitative and quantitative data through surveys and World Café sessions, collecting informed consent and sociodemographic details across all work packages to support the development and evaluation of an XR intervention. Work Package 1 gathers self-reported data on testicular cancer late effects, health status, care needs, and XR familiarity, while Work Package 2 produces evidence reviews, co-creation insights, and technical validation results. The subsequent phases (Work Packages 3–6) focus on generating data regarding the intervention's effectiveness on patient knowledge and quality of life, alongside comprehensive cost-	[Link]	[Link]

Project Acronym	Project Title	General Description of Data Generated	Link to Project-Specific DMP	Link to Project-Specific Data model
		effectiveness analyses, process evaluations, and assessments of ethical and policy implications.		

Sheet 2: Data Asset Inventory (The Core Sheet)

This is the most detailed sheet, designed to catalogue each distinct dataset and assess its readiness for sharing.

NOTE: so far is just a template to showcase how data will be feed during the WG-timeline.

Project	Data Asset ID	Data Type (Dropdown: Clinical, Genomic, Imaging, QoL Survey, etc.)	Brief Description	Format (e.g., CSV, VCF, DICOM)	Data Standard/Model Used (e.g., OMOP, FHIR, Custom)	Data Quality Procedure in Place? (Y/N) Justify which one or why not	Anonymisation Level (Dropdown: Anonymised, Pseudonymised)	Consent Allows Sharing with UNCAN.eu? (Y/N/Conditional)	Notes / Barriers
Project A	A-01	Clinical	Baseline patient demographics and treatment history.	CSV	OMOP-CDM v5.4	Y	Pseudonymised	Y	
Project A	A-02	QoL Survey	EORTC QLQ-C30 survey results.	REDCap CSV	Custom	Y	Pseudonymised	Conditional	Requires re-consent for federated analysis.
Project B	B-01	Genomic	Germline Whole Genome Sequencing data.	VCF	GA4GH	Y	Pseudonymised	Y	Data transfer requires secure enclave.

Sheet 3: Legal & Ethical Framework Comparison

This sheet helps identify commonalities and differences in the legal and ethical approaches across projects.

NOTE: so far is just a template to showcase how data will be feed during the WG-timeline.

Project name	Main contact person (name & Email address) for ethics & data protection for the consortium	Detail your Project's legal & ethics compliance process (how are you ensuring legal and ethical compliance?)	Where can the compliance process be found / Is specified? If found in a dedicated deliverable, can you share it with the cluster?	Did your project require a Data Protection Impact Assessment, a Fundamental Rights Impact Assessment or an Ethics Impact Assessment? I f so, which model/approach was used to perform them?	Did the project require signature of dedicated data sharing agreements / Data Processing Agreements? If so, please share the templates used with the cluster.	Did your project require ethical approvals? in which countries? If so, please share the approval templates used with the cluster

Find out more:

late-aya.eu

Co-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 European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.