Project Acronym:	HosmartAI
Grant Agreement number:	101016834 (H2020-DT-2020-1 – Innovation Action)
Project Full Title:	Hospital Smart development based on AI

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016834

DELIVERABLE

D6.7 – Data Management Handling Plan – First Version

Dissemination level:	PU -Public
Type of deliverable:	ORDP -Open Research Data Pilot
Contractual date of delivery:	30 June 2021
Deliverable leader:	EXYS
Status - version, date:	Final - v1.0, 2021-06-25
Keywords:	Data management handling plan, Data management strategy, Data management survey, Data storage, Data processing, FAIR, Quality control, Governance, Data security.

Executive Summary

This deliverable consists of the DMP (Data Management Handling Plan) for the HosmartAI project, which is funded by the European Union's H2020 Programme under Grant Agreement No. 101016834. The DMP aims to include among other (open) data sources, (open) source code, scientific publications, project deliverables and more. The DMP will be used internally by the consortium partners for the effective Data Management Handling (M2-M41, Leader: EXYS). Moreover, the DMP is a mandatory requirement for the HosmartAI project to participate in the Horizon 2020 Open Access to scientific peer-reviewed publications and research data.

The HosmartAI project participates in the Pilot on Horizon 2020 Open Research Data. The use of a Data Management Plan is required for all EU H2020 funded projects.

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to the project research data.

It also reflects the current state of the Consortium agreements on data management and must be consistent with exploitation and IPR requirements. Research data linked to exploitable results will not be put into the open domain if they compromise its commercialisation prospects or have inadequate protection, which is an H2020 obligation. The rest of the research data will be deposited in an open access repository.

The DMP covers the full data management life cycle for the data to be collected, processed and generated by the HosmartAI project. Towards handling research data management, within Publications and Research Data in Horizon 2020, detailing

1. how research data will be handled during & after the project;
2. what data will be collected, processed or generated;
3. what methodology & standards will be applied;
4. whether data will be shared /made open access/ how data will be curated and preserved.

The HosmartAI's DMP is based on the **FAIR principles**[1] (Findable, Accessible, Interoperable, Reusable)[2] and on the Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council (ERC) under Horizon 2020[3].

The Data Management Plan (DMP), which context is T6.5, will be formulated and delivered on M6 of the project in its First Version form. Since the DMP is expected to mature during the project, it will constitute a living document to be finalised and delivered as a final version at the end of the project. EXYS will lead the activity while all data provision partners will invest effort in safeguarding the proper data management.

[1] The FAIR Guiding Principles for scientific data management and stewardship, March 2016 https://www.nature.com/articles/sdata201618

[2] Guidelines Principles for FAIR, https://www.force11.org/fairprinciples

[3] https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

This first version of the DMP includes an overview of the datasets to be produced by the project, and the specific conditions that are attached to them. The next versions of the DMP will get into more detail and describe the practical data management procedures implemented by the project.

This DMP first evaluated the legal frameworks and the requirements of all pilots, then examined whether there are currently available data to which open access can be granted, always respecting the security and privacy requirements imposed. With regards to the dissemination of the scientific results, the consortium will establish and promote open access publications and partners will be encouraged to publish open access articles, so as to enable researchers to build upon previous research results, to foster collaboration, to avoid duplication of efforts, and to accelerate innovation.

Project members will be offered the option of publishing in journals contained/registered in the ROAR[4] and/or other repositories (OpenDOAR[5], OpenAIRE[6] and Zenodo[7]). Authors' copyright agreements will determine whether scientific publications, resulting from the project, will adopt the gold or the green model.

Among the task leader EXYS, contributors to this deliverable are **TGLV, PhE, UKCM, IRCCS, SERMAS, FIBHULP, CHUL, AHEPA, VUB, UM, INTRAS** and **PHILIPS.**

The structure of this DMP follows the H2020 Data Management Plan template.

Data Management Strategy:

To map all relevant data collections and to establish the data management needs, every WP was requested to complete a 'Data management survey' (Appendix B). The information collected there, are based on the following topics:

- Data summary and collection
- Data storage
- Data processing; cleaning, transforming and analysing
- Quality control
- Governance and security
- Ethical and legal issues

In addition to these tables, all WPs were allowed to insert additional comments related to data management which were not considered anywhere in the survey.
As it is described in chapter 1.1, the HosmartAI project will base its outcomes and improvements on eight large-scale pilots, therefore the research data management and the present DMP will be largely based on these pilots.

[4] ROAR: Research Open Access Repository, http://roar.eprints.org
[5] OpenDOAR: Directory for Open Access Repositories: https://v2.sherpa.ac.uk/opendoar/
[6] https://www.openaire.eu/
[7] https://zenodo.org /

Deliverable leader:	Angelo Consoli (EXYS)
Contributors:	Luca Gilardi, Angelo Consoli (EXYS) Georgios Rampidis (AHEPA, pilot #1) Nicola Bettin (VIMAR, pilot #3) SERMAS & FIBHULP (pilot #4) Marcela Chavez, Patrick Duflot (CHUL, pilot #2) Izidor Mlakar (UM, UKCM, pilot #5) Rosa Almeida (INTRAS, pilot #6) Robert Hofsink (PHILIPS, pilot #7) Magda Chatzikou (PhE) Dieter De Court, Wim Vranken (VUB, pilot #8) Carlos Luis Parra-Calderón (EFMI)
Reviewers:	Makis Karadimas (INTRA) Dieter De Court, Wim Vranken (VUB)
Approved by:	Athanasios Poulakidas (INTRA)

Document History

Version	Date	Contributor(s)	Description
0.1	2021-02-26	Luca Gilardi	Initial ToC
0.2	2021-03-01	Angelo Consoli	Review after internal EXYS meeting
0.3	2021-03-05	Luca Gilardi	Updated
0.4	2021-03-05	Angelo Consoli, Luca Gilardi	Added the DM Survey in section 1.1
0.5	2021-03-12	Angelo Consoli	Internal review before requesting partner contribution
0.6	2021-03-12	Luca Gilardi	Reference documents
0.7	2021-04-06	Luca Gilardi	Pilot nr. In survey tables.
0.8	2021-04-07	Izidor Mlakar	Added Data for Pilot #5
0.9	2021-04-13	Nicola Bettin	Added Data for Pilot #3
0.10	2021-04-15	Magda Chatzikou	Added Data for PhE
0.11	2021-05-26	Luca Gilardi	Live document with survey integration
0.12	2021-05-27	Marcela Chavez	Added data for CHUL
0.13	2021-05-27	Patrick Duflot	Added data for CHUL
0.14	2021-06-01	Carlos Luis Parra-Calderón	Added references to RDA FAIR Data Maturity Model and HL7 FHIR4FAIR IG
0.15	2021-06-03	Rosa Almeida	Added Data for Pilot #6
0.16	2021-06-03	Luca Gilardi	Freeze First Version's pre-release
0.17	2021-06-22	Luca Gilardi	Final version
1.0	2021-06-25	Irene Diamantopoulou, Athanasios Poulakidas	Final version for submission

Table of Contents

List of Tables

Definitions, Acronyms and Abbreviations

Acronym/ Abbreviation	Title
ASR	Automatic Speech Recognition
Cath lab	Catheterization laboratory
CCDS	Consensus Coding Sequence
CCTA	Coronary Computed Tomography Angiography
COBIT	Control Objectives for Information and related Technology
CPOE	Computerized Physician Order Entry
CRF	Case Report Form
CSV	Comma-Separated Values
CT	Computed Tomography
DICOM	Digital Imaging and COmmunications in Medicine
DMP	Data Management Plan
DPO	Data Protection Officer
DTA	Data Transfer Agreement
EHR	Electronic Health Record
ERC	European Research Council
FAIR	Findable, Accessible, Interoperable, Reusable
FFRCT	Fractional Flow Reserve derived from CT
FHIR	Fast Healthcare Interoperability Resources[8]
GDPR	General Data Protection Regulation[9]
HL7	Health Level Seven
iFR	instantaneous wave-Free ratio Resting index
IMU	Inertial Measurement Unit
IVUS	Intravascular Ultrasound
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LPD	Swiss Data Protection Law
OCT	Optical Coherence Tomography
PAM	Privilege Access Management
PGHD	Patient Generated Health Data
PREM	Patient Reported Experience Measure
PROM	Patient Reported Outcome Measure
RFR	Coronary Physiology Resting Full-Cycle Ratio
SOP	Standard Operating Procedure
SPSS	Statistical Package for the Social Sciences[10]
SSH	Secure Shell
SSL / TLS	Secure Sockets Layer / Transport Layer Security

[8] HL7 FHIR (Fast Healthcare Interoperability Resources), http://hl7.org/fhir/index.html
[9] https://gdpr-info.eu
[10] https://www.ibm.com/analytics/spss-statistics-software

Acronym/ Abbreviation	Title
SUS	System Usability Scale
TAM	Technology Acceptance Model
TBD	To be defined
TTS	Text-To-Speech
UEQ	User Experience Questionnaire
VPN	Virtual Private Network

Term	Definition
Accessible	Data is Accessible in that it can be always obtained by machines and humans upon appropriate authorization and through a well-defined protocol.
Cohort	In statistics, marketing and demography, a cohort is a group of subjects who share a defining characteristic (typically subjects who experienced a common event in a selected time period, such as birth or graduation).
Findable	Any Data Object should be uniquely and persistently identifiable
Interoperable	The ability of data or tools from non-cooperating resources to integrate or work together with minimal effort. Data Objects can be Interoperable only if: (Meta) data is machine-actionable, (Meta) data formats utilize shared vocabularies and/or ontologies, (Meta) data within the Data Object should thus be both syntactically parseable and semantically machine-accessible.
Pseudoanonymisation	The processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organizational measures to ensure non-attribution to an identified or identifiable individual[11]
Re-usable	For Data Objects to be Re-usable they should be sufficiently well-described and rich that it can be automatically (or with minimal human effort) linked or integrated, like-with-like, with other data sources. Published Data Objects should refer to their sources with rich enough metadata and provenance to enable proper citation.
Schemaless database	A type of database where each item is saved in its own document with a partial schema, leaving the raw information untouched.

[11] GDPR Article 4(3b): https://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm

1 Introduction

1.1 Project Information

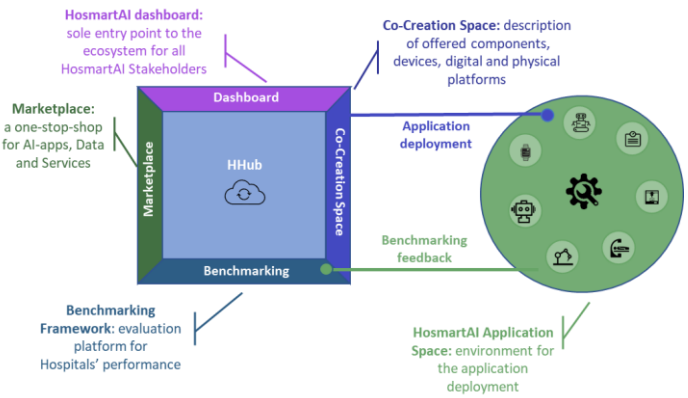
VISION

The HosmartAI vision is a strong, efficient, sustainable and resilient European **Healthcare system** benefiting from the capacities to generate impact of the technology European Stakeholders (SMEs, Research centres, Digital Hubs and Universities).

MISSION

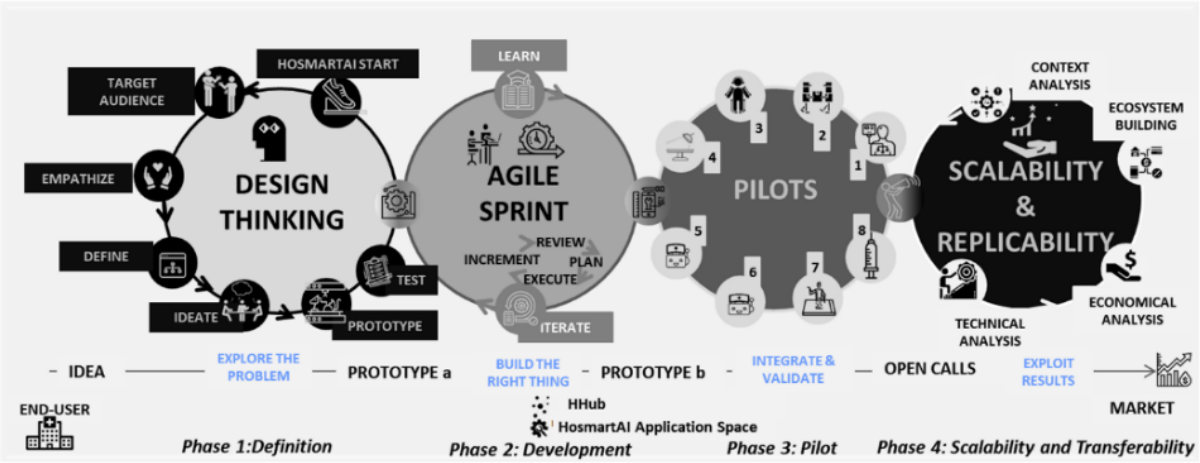
The HosmartAI mission is to guarantee the **integration** of Digital and Robot technologies in new Healthcare environments and the possibility to analyse their benefits by providing an **environment** where digital health care tool providers will be able to design and develop AI solutions as well as a space for the instantiation and deployment of a AI solutions.

HosmartAI will create a common open Integration **Platform** with the necessary tools to facilitate and measure the benefits of integrating digital technologies (robotics and AI) in the healthcare system.



A central **hub** will offer multifaceted lasting functionalities (Marketplace, Co-creation space, Benchmarking) to healthcare stakeholders, combined with a collection of methods, tools and solutions to integrate and deploy AI-enabled solutions. The **Benchmarking** tool will promote the adoption in new settings, while enabling a meeting place for technology providers and end-users.

Eight Large-Scale Pilots will implement and evaluate improvements in medical diagnosis, surgical interventions, prevention and treatment of diseases, and support for rehabilitation and long-term care in several Hospital and care settings. The project will target different **medical** aspects or manifestations such as Cancer (Pilot #1, #2 and #8); Gastrointestinal (GI) disorders (Pilot #1); Cardiovascular diseases (Pilot #1, #4, #5 and #7); Thoracic Disorders (Pilot #5); Neurological diseases (Pilot #3); Elderly Care and Neuropsychological Rehabilitation (Pilot #6); Fetal Growth Restriction (FGR) and Prematurity (Pilot #1).



To ensure a user-centred approach, harmonization in the process (e.g. regarding ethical aspects, standardization, and robustness both from a technical and social and healthcare perspective), the **living lab** methodology will be employed. HosmartAI will identify the appropriate instruments (**KPI**) that measure efficiency without undermining access or quality of care. Liaison and co-operation activities with relevant stakeholders and **open calls** will enable ecosystem building and industrial clustering.

HosmartAI brings together a **consortium** of leading organizations (3 large enterprises, 8 SMEs, 5 hospitals, 4 universities, 2 research centres and 2 associations – see Table 1) along with several more committed organizations (Letters of Support provided).

Table 1: The HosmartAI consortium.

Number[12]	Name	Short name
1 (CO)	INTRASOFT INTERNATIONAL SA	**INTRA**
1.1 (TP)	INTRASOFT INTERNATIONAL SA	**INTRA-LU**
2	PHILIPS MEDICAL SYSTEMS NEDERLAND BV	**PHILIPS**
3	VIMAR SPA	**VIMAR**
4	GREEN COMMUNICATIONS SAS	**GC**
5	TELEMATIC MEDICAL APPLICATIONS EMPORIA KAI ANAPTIXI PROIONTON TILIATRIKIS MONOPROSOPIKI ETAIRIA PERIORISMENIS EYTHINIS	**TMA**
6	ECLEXYS SAGL	**EXYS**
7	F6S NETWORK IRELAND LIMITED	**F6S**
7.1 (TP)	F6S NETWORK LIMITED	**F6S-UK**
8	PHARMECONS EASY ACCESS LTD	**PhE**
9	TERAGLOBUS LATVIA SIA	**TGLV**
10	NINETY ONE GMBH	**91**
11	EIT HEALTH GERMANY GMBH	**EIT**
12	UNIVERZITETNI KLINICNI CENTER MARIBOR	**UKCM**
13	SAN CAMILLO IRCCS SRL	**IRCCS**
14	SERVICIO MADRILENO DE SALUD	**SERMAS**

[12] CO: Coordinator. TP: linked third party.

Number[12]	Name	Short name
14.1 (TP)	FUNDACION PARA LA INVESTIGACION BIOMEDICA DEL HOSPITAL UNIVERSITARIO LA PAZ	**FIBHULP**
15	CENTRE HOSPITALIER UNIVERSITAIRE DE LIEGE	**CHUL**
16	PANEPISTIMIAKO GENIKO NOSOKOMEIO THESSALONIKIS AXEPA	**AHEPA**
17	VRIJE UNIVERSITEIT BRUSSEL	**VUB**
18	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	**AUTH**
19	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	**ETHZ**
20	UNIVERZA V MARIBORU	**UM**
21	INSTITUTO TECNOLÓGICO DE CASTILLA Y LEON	**ITCL**
22	FUNDACION INTRAS	**INTRAS**
23	ASSOCIATION EUROPEAN FEDERATION FORMEDICAL INFORMATICS	**EFMI**
24	FEDERATION EUROPEENNE DES HOPITAUX ET DES SOINS DE SANTE	**HOPE**

1.2 Document Scope

This deliverable aims at collecting all data that will be handled by consortium partners in the frame of the HosmartAI project.

The initial phase of data collection will be followed by the elaboration of the first release of this document, whose submission is planned at M6. This deadline is a mandatory requirement by the **Horizon 2020 Open Access Data Management** requirement: once a project has had its funding approved and has started, it must submit a first version of the DMP (as a deliverable) within the first 6 months of the project[13]. Other document updates are planned during the project duration; this will allow accommodating new findings and aligning the DMP to the needs of HosmartAI respecting active regulations.

1.3 Document Structure

This document is comprised of the following chapters:

Chapter 1 is an introduction to the project, to the document scope and structure, and gives an overview of the EU-H2020 Open Access programme and of the FAIR principles for data management applied to the HosmartAI project.

Chapter 2 concerns data summary, and presents the Data Management Survey, the project's pilots and purpose of the data collected, the datasets base information, the data types and formats, the physical location of the datasets, their expected sizes, as well as the purposes of the data (data utility) and identification.

[13] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Chapter 3 is based on the FAIR principles, and reports about findability of data, provisions for metadata, open accessibility, data interoperability, data re-use through licensing and quality assurance, as well as data cleansing, transforming and analysing.

Chapter 4 illustrates the allocation of resources for data management, i.e., costs for data FAIR and open access in the HosmartAI project, data management responsible, costs for long term preservation.

Chapter 5 concerns the provisions for data security and governance, reporting also the COBIT classification degree for every technical partner.

Chapter 6 presents the ethical and legal aspects linked to data management and data sharing.

Chapter 7 illustrates other issues about data management.

Chapter 8 presents tools and references involved in the management of data.

Chapter 9 lists the references of this document.

Chapter 10 presents the conclusions of this document.

Appendix A reports all data tables related to the datasets.

Appendix B presents the Data Management Survey filled-in by the pilots' leaders.

1.4 Open Access

The European Union (EU) strives to improve access to scientific information and to boost the benefits of public investment in the research funded under the EU Framework Programme for Research and Innovation Horizon 2020.

Launched by the European Commission along with the H2020 programme, **Open Access** is the practice of providing on-line access to scientific information that is free of charge to the reader and that is reusable. In the context of research and innovation, scientific information can refer to peer-reviewed scientific research articles or research data.

According to this strategy, in Horizon 2020 a limited pilot action on open access to research data has been implemented so that participating projects will be required to develop a Data Management Plan (DMP), in which they will specify what data will be open.

According to the H2020 **Article 29.2 of the Model Grant Agreement**[14], each beneficiary must ensure open access to all peer-reviewed scientific publications relating to its results. These open-access requirements are based on a balanced support to both 'Green open access' (immediate or delayed open access that is provided through self-archiving) and 'Gold open access' (immediate open access that is provided by a publisher). Apart from open access to publications, projects must also aim to deposit the research data needed to validate the results presented in the deposited scientific publications, known as "underlying data". In

[14] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf#page=242

order to effectively supply this data, projects need to consider at an early stage how they are going to manage and share the data they create or generate.

Nevertheless, data sharing in the open domain can be restricted as a legitimate reason to protect results that can reasonably be expected to be commercially or industrially exploited. Strategies to limit such restrictions will include anonymising or aggregating data, agreeing on a limited embargo period or publishing selected datasets.

From the Data Management point of view, Horizon 2020 strongly suggests that its beneficiaries would make their research data findable, accessible, interoperable and reusable **(FAIR)** to ensure that it is well managed. Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse. The HosmartAI's DMP follows the FAIR principles.

1.5 FAIR data concepts

The following two subsections report the FAIR main concepts as illustrated in the FORCE11's Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version B1.0[15]

By adopting all FAIR facets, Data Objects become fully: Findable, Accessible, Interoperable, and Reusable.

The FAIR Data principles aim to ensure that data are shared in a way that enables and enhances reuse by humans and machines. Although FAIR (Findable, Accessible, Interoperable, Reusable) emerged from a workshop for the life science community, the principles are intended to be applied to data and metadata from all disciplines. Since the formal release via the FORCE11 community, the FAIR data principles have been adopted by several funders and governments worldwide. The European Commission data management guidelines were updated in 2017 to introduce the notion of FAIR[16].

1.5.1 Definitions

- A **Concept** is any defined 'unit of thought' to which we refer in our digital formats.
- A **Data Object** is defined for the purpose of the principles below as: An **Identifiable Data Item with Data elements + Metadata + an Identifier**.
- When we use the term (**Meta) data** here, we intend to indicate that the principle is true for **Metadata** as well as for the actual, collected **Data Elements** in the Data Object, but that the principle in question can be independently implemented for each of them.

[15] Source : https://www.force11.org/fairprinciples

[16] Jaime Delgado et al., Approaches to the integration of TRUST and FAIR principles, section 1. Introduction, Universitat Politècnica de Catalunya et al.

1.5.2 FAIR Guiding Principles

1. To be **Findable** any Data Object should be uniquely and persistently identifiable.
   1.1. The same Data Object should be re-findable at any point in time; thus, Data Objects should be **persistent**, with emphasis on their metadata.
   1.2. A Data Object should minimally contain basic machine-actionable metadata that allows it to be distinguished from other Data Objects.
   1.3. Identifiers for any concept used in Data Objects should therefore be Unique and **Persistent**.
2. Data is **Accessible** in that it can be always obtained by machines and humans.
   2.1. Upon appropriate authorization.
   2.2. Through a well-defined protocol.
   2.3. Thus, machines and humans alike will be able to judge the actual accessibility of each Data Object.

3. Data Objects can be **Interoperable** only if:
   3.1. (Meta) data is machine-actionable.
   3.2. (Meta) data formats utilize shared vocabularies and/or ontologies.
   3.3. (Meta) data within the Data Object should thus be both syntactically parseable and semantically machine-accessible.

4. For Data Objects to be **Re-usable** additional criteria are:
   4.1. Data Objects should be compliant with principles 1-3.
   4.2. (Meta) data should be sufficiently well-described and rich that it can be automatically (or with minimal human effort) linked or integrated, like-with-like, with other data sources.
   4.3. Published Data Objects should refer to their sources with rich enough metadata and provenance to enable proper citation.

In order to know the maturity level of the application of FAIR principles in this DMP, each pilot will be checked using the specifications and guidelines of the Research Data Alliance (RDA) FAIR Data Maturity Model (https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0)

2 Data summary

As a first step, the following questions inherited from the indications of H2020 Data Management directives will be addressed:

- What is the purpose of the data collection/generation and its relation to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data?
- To whom might it be useful ('data utility')?

2.1 Data Management Survey

For the purpose of collecting information about the research data processed in the frame of the project, a Data Management Survey[17] was developed and proposed to the pilot leaders and technical partners. The involved partners are:

- **AHEPA** for pilot #1
- **CHUL** for pilot #2
- **IRCCS** for pilot #3
- **SERMAS and FIBHULP** for pilot #4
- **UM and UKCM** for pilot #5
- **INTRAS** for pilot #6
- **PHILIPS** for pilot #7
- **VUB** for pilot #8
- **TGLV** for various research data
- **PhE** for various research data

The information collected in the Survey was used to fill in this DMP.

Important notice:

This document is the first of a series of three deliverable documents D6.7, D6.8, D6.9 planned during the project. It is a living document and as such the same will be subject to update all during the HOSMARTAI project. D6.7 is based on a survey carried out during the first five months of the project. Seven pilots already contributed in this phase while pilot number #4 announced a small delay that will lead the pilot #4 leader to deliver its contribution after the date of submission of D6.7. Some aspects related to data management are still undefined open for many pilots and will be finalized in the next phases of the project for this reason this first version contains a certain number of “TBD” that will be clarified during the progress of the activities.

[17] The DMP Survey results are reported in the **Error! Reference source not found.** to this document

2.2 Purpose of the data (HosmartAI pilots)

As explained in the Introduction of this document (section 1.1), eight large-scale pilots will be implemented in the HosmartAI project, improving medical diagnosis, surgical interventions, prevention and treatment of diseases, and support for rehabilitation and long-term care in several hospitals and care settings. These pilots are targeting several **medical** aspects.

Table 2 gives an overview of the eight pilots, in relation to the technologies involved.

Table 2: The HosmartAI pilots.

Pilot #	Pilot title	Domain	Main technologies involved	Site	Pilot leader
1	Development of a clinician-friendly, interpretable computer-aided diagnosis system (ICADx) to support and optimise clinical decision making in multi-specialty healthcare environment.	Diagnosis Revolution	Computer-aided diagnosis system.	AHEPA Hospital & Hippokrateio General Hospital of Thessaloniki (Greece)	AHEPA
2	Optimizing the use of radiotherapy	Logistic Improvement	AI algorithm for optimizing patient scheduling.	CHUL Hospital (Belgium)	CHUL
3	Treatment Improvement with the use of innovative technologies and robotics in rehabilitation process	Treatment Improvement	Centralized data collection from wearable devices and environmental sensors.	IRCCS Rehabilitation Centre (Italy)	IRCCS
4	Robotic Systems for minimally Invasive Operation	Surgical support	Robotic system for cardiac catheter navigation, AI and Big Data techniques.	SERMAS Hospital (Spain)	SERMAS
5	Assistive Care in Hospital: Robotic Nurse	Assistive Care	Robotic nurse and Integration of data measured with digital devices.	UKCM Hospital (Slovenia)	UM
6	Assistive Care in Care Centre: Virtual Assistant	Assistive Care	Robotic nurse, Socially Assistive Robots, eCoach.	INTRAS Care Centre (Spain)	INTRAS

Pilot #	Pilot title	Domain	Main technologies involved	Site	Pilot leader
7	Smart Cathlab Assistant	Surgical Support	AI-enabled tools to provide real-time clinical decision support and to alleviate the administrative burden in the interventional suite.	UZ Brussel (Belgium)	PHILIPS/UZ Brussels
8	Prognosis of cancer patients and their response to treatment combining multi-omics data	Diagnosis and treatment Improvement	General framework to store and analyse raw medical data.	UZ Brussel (Belgium)	VUB

The main purpose of HosmartAI's pilots and the related data outcoming from the HosmartAI project is to improve efficiency in several areas of the medical field (as mentioned before). Moreover, data and datasets are intended to be made open to researchers in the field.

2.3 Datasets, base information

Base information of HosmartAI datasets consists of the type of data to be collected, the name of the datasets, the related pilots and tasks, as well as the responsible and collaborating partners in charge of managing the different datasets handled by the project.

Reference: Appendix A.1

2.4 Data types and formats, physical location

This section reports the actual physical location where the original datasets will be stored. Moreover, it gives the software tools used for data storage and the data standards used for storing the datasets.

Reference: Appendix A.2

2.5 Expected sizes and data volumes (nr. of records)

This section gives the expected size of the datasets used in the HosmartAI project.

Reference: Appendix A.3

2.6 Data utility and identification

Purpose of data and how they can be re-used (data utility) are important aspects of data management for this project. Reported here is the identifiability of collected information for each dataset and how the data will be made accessible to the consortium partners (and whether any restrictions apply).

Reference: Appendix A.4

3 FAIR

3.1 Making data findable, including provisions for metadata

This section shows if data are discoverable with metadata, identifiable and locatable by means of a standard identification mechanism, reports the identified naming convention, and the search keywords (if any) and datasets version number that could optimize the re-use.

Reference: Appendix A.5

3.2 Making data openly accessible

All aspects of data open accessibility are covered in this section. This includes which datasets are made available openly and on what open repositories those will be hosted, as well as the licenses accompanying them, the access identifications and the possible restrictions.

Reference: Appendix A.6

3.3 Making data interoperable

To identify interoperability of the HosmartAI datasets, metadata vocabularies, ontologies, standards and general methodologies for data interoperability are provided.

Reference: Appendix A.7

3.4 Increase data re-use (through clarifying licenses)

Re-use of data from the project's datasets is accomplished by answering the following questions:

- How will the data be licensed to permit the widest re-use possible?
- When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.
- Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.
- How long is it intended that the data remains re-usable?
- Are data quality assurance processes described?

3.4.1 Data licensing, availability and usability by third parties

This sub-section reports the availability, usability and licensing of the project data towards third parties.

Reference: Appendix A.8

3.4.2 Data Quality Assurance processes

Data Quality Assurance is achieved by specifying for each pilot and task number/dataset the types of analysis that will be performed, the person in charge to create the statistical analysis plan, and how the transformations and analyses on the data will be verified.

Reference: Appendix A.9

3.4.3 Data cleansing, transforming and analysing

Several post-processing activities are envisaged on the project's datasets. Information about data cleansing, transforming and analysing include the type of data cleaning needed (e.g., correct data types, remove duplicates, add missing info…), the person responsible for data cleaning, the type of data transformation/analysis (e.g., normalization, discretization, …), the software/tools used for cleaning, transform, and analyse, where and by whom will the analysis be conducted, finally the standards followed for code development / access and re-use (if available).

Reference: Appendix A.10

4 Allocation of resources

This section reports the expected costs for making data FAIR in the HosmartAI project and for long term preservation, and how these costs will be covered (taking into consideration that the costs related to open access to research data are eligible as part of the Horizon 2020 grant). The responsible for data management is also indicated, when applicable.

4.1 Costs of FAIR and non-FAIR data in HosmartAI

Table 3: Costs for making data FAIR

Pilot #	Expected costs for making data FAIR
1	TBD
2	MosaiQ Machine. The relevant data are free by DICOM port for ELEKTRA. Electronic Health Record. History of patient (OMNIPRO).
3	TBD
4	TBD
5	Retrospective data will not be shared. The cost of transforming data to FAIR would include: 1. Extraction of data from the specific clinical cohorts (manual extraction via queries) 2. Transformation to FHIR model (automatic, proprietary software) 3. Iterative process(half-automated): Verification of level of anonymity and anonymization (i.e., removal of risks related to re-identification). Open-source tools can be used to decrease the cost. For prospective data: 1. Extraction of data from the specific clinical cohorts (automatic, since prospective Patient generated health data – PGHD - will be stored in FHIR) 2. Iterative process (half-automated): Verification of level of anonymity (half-automated) and anonymization (i.e., removal of risks related to re-identification). Open-source tools can be used to decrease the cost.
6	TBD
7	TBD
8	TBD

4.2 Responsible for data management

Table 4: Data management responsible.

Pilot #	Data management responsible
1	TBD
2	TMA: Collect and store data from machine. CHUL: Generate anonymized datasheet.
3	TBD
4	TBD

5	Patient is the data owner, UKCM is data controller, UM, ITCL and GC are data processors. UKCM retrieves patient consent, exports the data and defines the conditions regarding the level of anonymization and use. UKCM anonymizes data and ensures the level of anonymization is maintained.
6	TBD
7	TBD
8	TBD

4.3 Costs for long term preservations

Table 5: Long term preservation costs.

Pilot #	Long term preservation costs
1	TBD
2	CHUL hospital have to analyse if the management of datasheet anonymized have cost for long term preservations.
3	TBD
4	TBD
5	The process of exports of anonymized and full de-identifiable data is done per request and is not stored separately by the UKCM. In HosmartAI the PGHD (pseudo anonymized and fully identifiable) will be stored on site, on a dedicated FHIR server in a private area network (PAN).
6	TBD
7	TBD
8	TBD

5 Data security

This section aims at answering the following questions:

- What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?
- Is the data safely stored in certified repositories for long term preservation and curation?

5.1 Provisions for data security and governance

Classification

For the assessment of governance and security, the maturity of the process is classified in maturity levels using values that are inherited from the COBIT standardization guidelines[18]. Those values are:

Table 6: Reports the COBIT classification degree on data security provision for every technical partner.

Acronym	Name	Description
I	**Initial**	Process not specified, based on spontaneous initiative, that is poorly controlled and reactive.
M	**Managed**	Process is planned, documented and monitored at the project level but not integrated in a broader scope at organization level.
D	**Defined**	Proactive process active at organization level.
Q	**Quantitatively Managed**	The process is measured and controlled / Verified.
O	**Optimizing**	Focus is on continuous process and improvement.

Table 7: Data security.

Partner short name	Classification	Code of conduct used for each task	Levels of data security in place.	Levels of security that primary data used in the project will undergo	Name and email of the data privacy officer(s)
EXYS	Q	Data handling procedures at EXYS are defined according to GDPR and LPD (Swiss data protection law)	Technical measures: • Data storage in dedicated servers segmented on network environment • Authorized access with data access audit logs • VPN regulated access to the data processing data centre	Whenever required by the project, data will be stored in firewall-protected computers with strong authentication.	Angelo Consoli; <angelo.consoli@eclexys.com>

[18] ISACA®, *COBIT® 2019 Framework: Governance and Management Objectives*, USA 2018

Partner short name	Classification	Code of conduct used for each task	Levels of data security in place.	Levels of security that primary data used in the project will undergo	Name and email of the data privacy officer(s)
			• Data communication and transfer protected by Secure Socket Layer (SSL) and Transport Layer Security (TLS) protocols.		
AUTH	N/A	Data are handled according to GDPR and international ethical guidelines (inter alia, the Word Medical Association Declaration of Helsinki), mandated by the AUTH Research Ethics Committee. Ethical approval must be obtained before data collection involving human beings.	Technical measures include: • Data storage in firewall-protected computers • Authorized access with data access audit logs • Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols.	• Data will be stored in firewall-protected computers with authorized access. • Access will be limited to members of the AUTH research team. • In case of data transfer, this will take place via SSH or SSL protocols. • Pseudo-anonymised data will be stored separately from records including subjects' personally identifiable information (e.g., signed consent forms)	Ms. Kornilia Skarpeta <data.protection@auth.gr>
VIMAR	M	Data are handled according to GDPR.	Technical measures include: • Data storage in firewall-protected computers. • Authorized access with data access audit logs. • Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols.	• Data will be stored in firewall-protected computers with authorized access. • Access will be limited to authorized members of the team. • In case of data transfer, this will take place via SSL protocols. • Pseudo-anonymised data will be stored separately from records including subjects' personally	Beni Luigi Gianesin <beni.gianesin@vimar.com>

Partner short name	Classification	Code of conduct used for each task	Levels of data security in place.	Levels of security that primary data used in the project will undergo	Name and email of the data privacy officer(s)
				identifiable information.	
PhE	N/A	Data analysis will be handled according to GDPR following PhE Standard Operating Procedure of Process & Handling of Personal Data.	Cloud – One Drive	Data will be password protected and accessed only by the PhE HOSMARTAI dedicated team.	George Tsonis <Tsonislaw@yahoo.gr>
CHUL	Q	Data are handled according to GDPR mandated by both DPO and the CHU Liège Ethics Committee			Ghislaine.Dumont@chuliege.be
INTRAS	Q	Data handling procedures at INTRAS are defined according to GDPR	Technical measures: - Data storage in dedicated servers segmented on network environment - Authorized access with data access audit logs - VPN regulated access to the data processing data centre Data communication and transfer protected by Secure Socket Layer (SSL) and Transport Layer Security (TLS) protocols.	- Data will be stored in firewall-protected computers with authorized access. - Access will be limited to members of the INTRAS research team. - In case of data transfer, this will take place via SSH or SSL protocols. Pseudo-anonymised data will be stored separately from records including subjects' personally identifiable information (e.g., signed consent forms)	Francisco Mallo Vázquez <fmv@intras.es>
UM	N/A	Data are handled according UM's privacy policy (https://feri.um.si/en/about-us/privacy-policy/)	Technical measures include: • Data storage in firewall-protected computers	• UM will not store sensitive data	Doc. dr. Miha Dvojmoč (dpo@um.si), if contacted please always refer project and the main contact person of the project

Partner short name	Classification	Code of conduct used for each task	Levels of data security in place.	Levels of security that primary data used in the project will undergo	Name and email of the data privacy officer(s)
			• Authorized access with data access audit logs • Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols.		(izidor.mlakar@um.si)
UM	N/A	Data are handled according UM's privacy policy (https://feri.um.si/en/about-us/privacy-policy/)	Technical measures include: • Data storage in firewall-protected computers • Authorized access with data access audit logs • Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols.	• UM will not store sensitive data	Doc. dr. Miha Dvojmoč (dpo@um.si), if contacted please always refer project and the main contact person of the project (izidor.mlakar@um.si)
UKCM	M	Data are handled according to GDPR and international ethical guidelines (inter alia, the Word Medical Association Declaration of Helsinki), mandated by the UKCM's and National Ethics Committee. Ethical approval must be obtained before data collection involving human beings.	Technical measures include: • Data storage in firewall-protected computers • Authorized access with data access audit logs • Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols.	• Data will be stored in firewall-protected computers with authorized access. • Access will be limited to members of the UKCM and UM research team. A specific DPA will be established. • In case of data transfer, this will take place via SSH or SSL protocols. • Pseudo-anonymised data will be stored separately from records including subjects' personally identifiable information (e.g., signed consent forms)	mag. Klara Mihaldinec, (dpo@ukc-mb.si)

6 Ethical and legal aspects

Ethical and legal aspects which are part of the HosmartAI project are treated in the present section. The questions to which an answer is provided are:

- Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).
- Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

6.1 Ethical issues for data sharing

Data Protection is a fundamental human right, enshrined in the EU Charter of Fundamental Rights, aimed at providing any individual[19] with control over the way information about him/her is collected and used. Article 8(1) of the Charter of Fundamental Rights of the European Union (the 'Charter') and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU) grant everyone the right to the protection of personal data concerning him or her. Data protection is a central issue for research ethics.

Whenever personal data is collected, there are both ethical and legal obligations to ensure that participants' information is properly protected. This is fundamental to safeguarding their rights and freedoms, and minimising the ethics risks related to the data processing. In HosmartAI data security is provided on all levels. Only authorized users will have access to digital information. The Project will adopt recommendations and standards provided by ENISA (European Union Agency for Cybersecurity). It is the goal of all project partners to mitigate the risk for all participating patients

Publication of Results HosmartAI complies with the highest ethical standards. Researchers, authors, sponsors, editors and publishers all have ethical obligations with regard to the publication and dissemination of the results of research.

6.1.1 Ethical review

This sub-section is dealing with the type of ethical review needed for each pilot

Reference: Appendix A.11

6.2 Legal issues for data sharing

The GDPR provides the basic legal framework for personal data processing and therefore data sharing. Particular attention has to be paid to research involving sensitive data such as health data, which according to GDPR must not be processed unless the data subject has given

[19] An individual is an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person (Art. 2(a) EU General Data Protection Regulation (GDPR).

explicit consent. This imposes obligations on researchers to provide research subjects with detailed information about what will happen to the personal data that they collect.

In HosmartAI all data processing will comply with EU law as well as national data laws. It will be ensured that any partners, contractors or service providers that process research data at HosmartAI partners' request and on their behalf will comply with the GDPR and the H2020 ethics standards. Special attention will be given to a good balance between research objectives and the means used to achieve them.

Types of Personal Data Processed During the lifetime of HosmartAI two categories of personal and sensitive data are/may be generated or collected: (1) "Data related to stakeholders" (individuals working for the Consortium partners or in any way professionally involved with the Project, etc.): Information on these data subjects such as contact details (like e-mails and names), their signatures, authorship of deliverables, etc. is collected and processed by all partners of the Project. (2) "Patient's health data in pilots" (sensitive data pr "special personal data" pursuant to Article 9 GDPR): At this stage of the project, it has not been decided which criteria may be introduced to limit or exclude from the data processing patients on an individual basis.

Informed Consent When personal data is used, informed consent is the cornerstone of research ethics. The lawful basis for the processing of personal data related to stakeholders, under the GDPR, is that each data subject working for a Project partner has given consent to the processing of his or her personal data (GDPR Article 6 (1)(a)) and that the processing is necessary for the performance of a contract - namely, the data subjects' employment agreements with each Project partner (GDPR Article 6(1)(b)). After the end of the project, files containing personal information of data subjects working for Project partners will be maintained by each Project partner. Any partner will have the right to continue to maintain its copy of the contact data of employees working for HosmartAI partners unless said that employees request a deletion of the contact data. Mailing lists of the project will be deleted only after the very final payment and assessment from the European Commission. Data subjects' contact details will be shared only with Project members and only for the time needed to execute the Grant Agreement and/or complete the Project. Authorship information may be made publicly available with the consent of the data subjects once the application becomes publicly or commercially available.

Whenever personal data is collected from patients, the patients' informed consent must be sought by means of a procedure that meets the minimum standards of the GDPR. This requires consent to be given by a clear affirmative act. For consent to data processing to be 'informed', the data subject must be provided with detailed information about the envisaged data processing in an intelligible and easily accessible form, using clear and plain language. The researchers at the Pilots will explain to patients (i)what the research is about; (ii)what their participation in the project will entail and in what risks they may be involved. The partner will give information as to whether data will be shared with or transferred to third parties and for what purposes and how long the data will be retained before they are destroyed. The patients will be also be informed about the right to withdraw consent or access to their data.

They will also be told the procedures to follow should they wish to do so. They will also receive information on their right to lodge a complaint with a supervisory authority. The data subjects must also be made aware if data are to be used for any other purposes, and if it is shared with research partners or transferred to organisations outside the EU. Records documenting the informed consent procedure, including the information sheets and consent forms provided to research participants, will be kept. The consent process(es) and the information provided to the data subjects will cover all the data processing activities related to their participation in HosmartAI. If in the course of the HosmartAI research project any significant changes to methodology or processing arrangements that have a bearing on the data subjects' rights or the use of their data, should occur, the data subjects will be made aware of the intended changes, and their express consent further use of the data will be sought.

Privacy by Design To innovate ethically and responsibly, researchers and developers apply the concept of 'privacy by design', which provides a framework for focusing the design of systems, databases and processes on respect for data subjects' fundamental rights. A wider concept of 'data protection by design', now included in the GDPR, requires the implementation of appropriate technical and organisational measures to give effect to the GDPR's core data-protection principles. Data protection by design is one of the best ways to address the ethics concerns that arise within a research Project. Minimisation of data is essential in this respect. Data processing must be lawful, fair and transparent. It should involve only data that are necessary and proportionate to achieve the specific task or purpose for which they are collected.

Deletion and Archiving of Data Personal data will only be kept as long as it is necessary for the purposes for which they are collected, or in accordance with the established auditing, archiving or retention provisions of HosmartAI. As soon as the research data is no longer needed, or subject to an established retention period, the data will be deleted. Data retained for auditing processes will be stored securely and further processed for those purposes only. Research data held in the cloud or by a third-party service provider, will also be together with any back-ups.

Reuse of Data A potential later use of the HosmartAI platform may permit medical researchers to use data sets for the purpose of conducting medical research. The procedure for this possibility has not yet been addressed by the Project partners. As a result of this effort, ethical and legal considerations may arise with respect to large scale or big data processing and they will be discussed in a later version of this DMP.

7 Other issues about data management

This section reports on possible other issues related to data management in the HosmartAI project. For instance, the use of other national/funder/sectorial/departmental procedures for data management.

Pilot #	Other issues related to data management
1	TBD
2	CHUL hospital have to do the creation and management of datasheet anonymized.
3	TBD
4	TBD
5	UKCM prefers the decentralized approach in which data is stored within the Pilot site. If sharing of prospective data established UKCM must create and anonymize data. Individual DPAs with clearly identified intents of the use of data and how data will be handled must be negotiated.
6	TBD
7	TBD
8	TBD

8 Tools and references

- The Research Data Alliance provides a Metadata Standards Directory that can be searched for discipline-specific standards and associated tools.
- Research Data Alliance FAIR Data Maturity Model (https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0).
- The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.
- Useful listings of repositories include:
  - Registry of Research Data Repositories
  - Some repositories like Zenodo, an OpenAIRE and CERN collaboration), allow researchers to deposit both publications and data, while providing tools to link them.
- Other useful tools include DMP online and platforms for making individual scientific observations available such as ScienceMatters.
- Mosaiq Machine data for radiotherapy by FHIR.
- Omnipro[20]. IT CHUL hospital transfer the anonymized therapy data by FHIR.
- FHIR4FAIR Implementation Guide (HL7 project underway, ballot scheduled for September 2021) (http://build.fhir.org/ig/HL7/fhir-forfair/)
- ROAR: Research Open Access Repository, http://roar.eprints.org
- OpenDOAR: Directory for Open Access Repositories: https://v2.sherpa.ac.uk/opendoar/
- OpenAIRE: https://www.openaire.eu/
- Zenodo: https://zenodo.org/
- Elekta MosaiQ Radiation Oncology: https://www.elekta.com/software-solutions/care-management/mosaiq-radiation-oncology/
- GRADIOR: Computer-based cognitive rehabilitation program.
- EVA Corpus: A Corpus for Analysing Linguistic and Paralinguistic Features in Multi-Speaker Spontaneous Conversations
- HBASE, MONGO
- K-Anonymity: https://github.com/Nuclearstar/K-Anonymity
- ARX - Open Source Data Anonymization Software: https://github.com/arx-deidentifier/arx
- European open Science Cloud (EOSC): https://eosc-portal.eu/

[20] https://cabinetprive.xperthis.com/omnipro/

9 References

[REF-01]	AGA - Annotated Model Grant Agreement, H2020 Programme, v. 5.2., June 2019
[REF-02]	Template for the Data Management Plan, H2020 Programme
[REF-03]	FAIR Data Management template Summary table, H2020 programme
[REF-04]	H2020 Model Grant Agreement - Article 29.2
[REF-05]	FORCE11's Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing (https://www.force11.org/fairprinciples)
[REF-06]	ENISA: European Union Agency for Cybersecurity (https://www.enisa.europa.eu/)
[REF-07]	Jaime Delgado, FAIR4Health - Report on Security and Privacy in FAIR processes, Horizon 2020 project, grant agreement No 824666
[REF-08]	Jaime Delgado and al., Approaches to the integration of TRUST and FAIR principles, Universitat Politècnica de Catalunya (UPC BarcelonaTECH)

10 Conclusions

This first version of the DMP includes an overview of the datasets to be produced by the project, and the specific conditions that are attached to them. Information was collected by proposing a Survey to the involved partners (both the pilot leaders and the other technical partners). As such, many information is still missing. The gap will be bridged in the following two versions (deliverables D6.8 and D6.9)

Integration of Future Aspects: the DMP is a living document and further considerations will be made, especially with respect to donated health records and potential for the application to be used as a research platform, in later iterations. Some aspects related to data management are still undefined open for many pilots and will be finalized in the next phases of the project for this reason this first version contains a certain number of “TBD” that will be clarified during the progress of the activities.

Appendix A Datasets

A.1 Datasets base information

Table 8: Datasets base information.

Code	Type of data to be collected / name of the dataset	Pilot Nr.	Task Nr.	Responsible partner	Collaborating partners
DS1.1	Cardiac ultrasound video recordings	1	3.1, 5.2	AHEPA	AUTH
DS1.2	Capsule endoscopy video recordings	1	3.1, 5.2	AHEPA	AUTH
DS1.3	Cardiotocography variables and results, biometric data, medical history data	1	3.1, 5.2	AUTH	N/A
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	1	3.1, 5.2	AHEPA	AUTH
DS2.1	Data related to hours spent by specialists.	2	3.2	CHUL	ITCL/TM
DS2.2	Retrospective patient schedule data and precondition of the treatment	2	3.2	CHUL	ITCL/TMA
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	2	1.2	CHUL	N/A
DS2.4	PROMs/PREMs	2	5.2	CHUL	UM
DS2.5	Prospective patient clinical data	2	5.2	CHUL	ICTL/TMA
DS2.6	Patient personal data (address, preferences, …)	2	5.2	CHUL	ICTL/TMA
DS2.7	Retrospective EHR	2	3.2	CHUL	ICTL/TMA
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction	2	3.2	ITCL	ICTL/TMA
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption logs, Human	3	3.3	IRCCS	VIMAR

Code	Type of data to be collected / name of the dataset	Pilot Nr.	Task Nr.	Responsible partner	Collaborating partners
	presence/access control, Devices' activation and connected loads				
DS3.2	IMU data captured by iPrognosis smartphone application	3, 6	5.2	AUTH	INTRAS
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	3, 6	5.2	AUTH	INTRAS
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	3, 6	5.2	AUTH	INTRAS
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	3, 6	5.2	AUTH	INTRAS
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	6	3.5	INTRAS	ITCL/AUTH
DS4.xx					
DS5.1	Audio-Visual recordings of patients	5	5.2	UM	UKCM, ITCL
DS5.2	Biometric data (e.g., blood pressure and heart rate)	5	5.2	UKCM	UM
DS5.3	Retrospective electronic health records	5	5.2	UKCM	UM
DS5.4	PREMs related to clinical staff (depends on T1.4)	5	5.2	UKCM	UM
DS5.5	PREMs related patients (PAM, SUS-SI/TAM, UEQ)	5	5.2	UKCM	UM
DS5.6	Datasets for facial expression and emotion recognition	5	3.5	UM	N/A
DS5.7	Datasets for ASR and TTS in Slovenian	5	3.5	UM	N/A
DS5.8	Datasets for ASR and TTS in French	5	3.5	UM	N/A

Code	Type of data to be collected / name of the dataset	Pilot Nr.	Task Nr.	Responsible partner	Collaborating partners
DS5.9	EVA Corpus[21], data set of conversational expression	5	3.5	UM	N/A
DS5.10	Video recordings of third persons	5	3.5	ITCL	N/A
DS5.11	Patient's behavioural information	5	3.5	ITCL	UM
DS5.12	Patient's facial information	5	3.5	ITCL, UM	N/A
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data.	6	3.5	ITCL	N/A
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	7	3.6	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	7	3.6	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	7	3.6	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	7	3.6	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique	7	3.6	Cardiology department of University	Philips Image Guided Therapy

[21] https://www.iaras.org/iaras/home/cijc/a-corpus-for-analyzing-linguistic-and-paralinguistic-features-in-multi-speaker-spontaneous-conversations-eva-corpus

Code	Type of data to be collected / name of the dataset	Pilot Nr.	Task Nr.	Responsible partner	Collaborating partners
	before and after a coronary intervention			Hospital Brussels, VUB	Systems (Philips)
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	7	3.6	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	7	3.6	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)
DS8.1	Image, gene, phenotype and pathology data for glioma patients	8	TBD	VUB	N/A
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	PhE	5.3	All pilots	PhE

A.2 Data types and formats, physical location

Table 9: Data types, formats and physical location.

Code	Type of data to be collected / name of the dataset	Physical location where primary (original) data is stored	Software/ tools used for storage of data?	Format and type of data standards used to store the data
DS1.1	Cardiac ultrasound video recordings	At the edge, dedicated infrastructure for the pilot	TBD	DICOM, HL7-compatible resource
DS1.2	Capsule endoscopy video recordings	At the edge, dedicated infrastructure for the pilot	TBD	MP4, HL7-compatible resource
DS1.3	Cardiotocography variables and results, biometric data, medical history data	At the edge, dedicated infrastructure for the pilot	TBD	TBD

Code	Type of data to be collected / name of the dataset	Physical location where primary (original) data is stored	Software/ tools used for storage of data?	Format and type of data standards used to store the data
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	At the edge, dedicated infrastructure for the pilot	TBD	DICOM partially (others TB
DS2.1	Data related to hours spent by specialists.	CHUL	N/A	HL7-compatible resource
DS2.2	Retrospective patient schedule data and precondition of the treatment	CHUL	Mosaiq OIS	HL7-compatible resource
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	CHUL	Mosaiq OIS	DICOM, HL7-compatible resource
DS2.4	PROMs/PREMs	CHUL	CHUL	TBD
DS2.5	Prospective patient clinical data	CHUL	EHR	HL7-compatible resource.
DS2.6	Retrospective EHR	CHUL	EHR	TBD
DS2.7	Data from radiotherapy services, infrastructure. Patient's satisfaction, preferences....	At the edge, dedicated infrastructure for the pilot	Data Store (HBASE, MONGO, RDBMS)	Media Resource and formats of HBASE, MONGO, RDBMS observations
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption logs, Human presence/access control, Devices	On VIMAR cloud infrastructure	InfluxDB	JSON

Code	Type of data to be collected / name of the dataset	Physical location where primary (original) data is stored	Software/ tools used for storage of data?	Format and type of data standards used to store the data
	activation and connected loads			
DS3.2	IMU data captured by iPrognosis smartphone application	Cloud infrastructure for the pilots	Schemaless database	JSON, HL7-compatible resource
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	Cloud infrastructure for the pilots	Schemaless database	JSON, HL7-compatible resource
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	Cloud infrastructure for the pilots	Schemaless database	JSON, HL7-compatible resource
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	Cloud infrastructure for the pilots	Schemaless database	JSON, HL7-compatible resource
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	Cloud infrastructure for the pilots	TBD	JSON
DS4.xx	TBD in D6.8			
DS5.1	Audio-Visual recordings of patients	At the edge, dedicated infrastructure for the pilot	Open FHIR Server	Media Resource of FHIR Observation
DS5.2	Biometric data (e.g., blood pressure and heart rate)	At the edge, dedicated infrastructure for the pilot	Open FHIR Server	FHIR Diagnostic Report, and Observation resource
DS5.3	Retrospective electronic health records	At the edge, interface with existing IT system of the hospital	Proprietary Hl7 Compliant Medis Server	N/A

Code	Type of data to be collected / name of the dataset	Physical location where primary (original) data is stored	Software/ tools used for storage of data?	Format and type of data standards used to store the data
DS5.4	PREMs related to clinical staff (depends on T1.4)	A GDPR compliant online survey infrastructure (https://1ka.arnes.si/index.php?lang_id=2)	SPSS or similar	JSON/CSV
DS5.5	PREMs related patients (PAM, SUS-SI/TAM, UEQ)	A GDPR compliant online survey infrastructure (https://1ka.arnes.si/index.php?lang_id=2)	SPSS or similar	JSON/CSV
DS5.6	Datasets for facial expression and emotion recognition	At UM's internal infrastructure	N/A	N/A
DS5.7	Datasets for ASR and TTS in Slovenian	At UM's internal infrastructure	N/A	N/A
DS5.8	Datasets for ASR and TTS in French	At UM's internal infrastructure	N/A	N/A
DS5.9	EVA Corpus, data set of conversational expression	At UM's internal infrastructure And CLARIN.SI Repository	N/A	N/A
DS5.10	Video recordings of third persons	Inside the robot hardware or attached devices	N/A	Bytestreams
DS5.11	Patient's behavioural information	At the edge, dedicated infrastructure for the pilot	Data Store (HBASE, MONGO, RDBMS)	To be defined
DS5.12	Patient's facial information	At the edge, dedicated infrastructure for the pilot	Data Store (HBASE, MONGO, RDBMS)	FAPS, AUs

Code	Type of data to be collected / name of the dataset	Physical location where primary (original) data is stored	Software/ tools used for storage of data?	Format and type of data standards used to store the data
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data	At the edge, dedicated infrastructure for the pilot	Data Store (HBASE, MONGO, RDBMS)	Media Resource and formats of HBASE, MONGO, RDBMS observations
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	UZB, VUB	Electronic health records system branded PRIMUZ	xlsx files
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	UZB, VUB	Electronic health records system branded PRIMUZ	xlsx files
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	UZB, VUB	Philips cathlabs over the Philips IntelliSpace CardioVascual (ISCV) Portal	DICOM
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	UZB, VUB	Coroventis platform (RFR/FFR) and Volcano platform (iFR/FFR)	Data extracted in .dat or .xls files
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	UZB, VUB	Abbott OPTIS platform (OCT) (Integrated platform) Volcano platform	TBD

Code	Type of data to be collected / name of the dataset	Physical location where primary (original) data is stored	Software/ tools used for storage of data?	Format and type of data standards used to store the data
			(IVUS) (non-integrated platform)	
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	UZB, VUB	Philips (CT) Heartflow (FFRCT)	TBD
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	Philips AI smart-cathlab prototype working at UZB, VUB	TBD	TBD
DS8.1	Image, gene, phenotype and pathology data for glioma patients	Separate in-hospital databases, with final central integration on site	PRIMUZ, XNAT	RDMS (various), JSON
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	The original data at each pilot's repository. The analysis of PhE will be stored in cloud space of PhE (One Drive)	R, Stata	Econometric data, *dta, *xlsx,

A.3 Expected data sizes and volumes

Table 10: Datasets expected size.

Code	Type of data to be collected / name of the dataset	Record size and expected data volume (n. of records)
DS1.1	Cardiac ultrasound video recordings	TBD
DS1.2	Capsule endoscopy video recordings	Recordings from 60 patients

Code	Type of data to be collected / name of the dataset	Record size and expected data volume (n. of records)
DS1.3	Cardiotocography variables and results, biometric data, medical history data	TBD
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	TBD
DS2.1	Data related to hours spent by specialists.	TBD
DS2.2	Retrospective patient schedule data and precondition of the treatment	2500 patients/year
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	Data of 5 machines
DS2.4	PROMs/PREMs	TBD
DS2.5	Prospective patient clinical data	TBD
DS2.6	Patient personal data (address, preferences, …)	TBD
DS2.7	Retrospective EHR	Subset to be identified if needed
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction	TBD
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption logs, Human presence/access control, Devices' activation and connected loads	TBD
DS3.2	IMU data captured by iPrognosis smartphone application	TBD
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	TBD
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	TBD
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	TBD
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	TBD
DS4.xx		

Code	Type of data to be collected / name of the dataset	Record size and expected data volume (n. of records)
DS5.1	Audio-Visual recordings of patients	1 recording per interaction, assuming 5-10minute interaction up to 300 MB per recording
DS5.2	Biometric data (e.g., blood pressure and heart rate)	TBD
DS5.3	Retrospective electronic health records	Up to 100MB per patient
DS5.4	PREMs related to clinical staff (depends on T1.4)	Up to 2 MB per patient
DS5.5	PREMs related patients (e.g., PAM, SUS-SI/TAM, UEQ) and PROs	Up to 2 MB per patient
DS5.6	Datasets for facial expression and emotion recognition	➢ 10GB
DS5.7	Datasets for ASR and TTS in Slovenian	TBD we estimate at least 20h of speech
DS5.8	Datasets for ASR and TTS in French	TBD we estimate at least 20h of speech
DS5.9	EVA Corpus, data set of conversational expression	2GB per annotated 1h recording
DS5.10	Video recordings of third persons	TBD
DS5.11	Patient's behavioural information	TBD
DS5.12	Patient's facial information	TBD
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data.	At least 1 recording per day per patient and per signal
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Around 1100 coronary angiogram including 350 CA with PCI performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Around 1100 coronary angiogram including 350 CA with PCI performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	Around 1100 coronary angiogram including 350 CA with PCI performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021

Code	Type of data to be collected / name of the dataset	Record size and expected data volume (n. of records)
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	Around 150 single point and 20 motorized pullbacks of physiological evaluations by FFR/iFR/RFR performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	Around 20 OCT and 10 IVUS performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	Around 300 CTA performed in patient referred for an invasive angiogram performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	Validation study of the AI prototype effects on key performance indicators (hospital and health care productivity). Prospective cohort of 100 patients.
DS8.1	Image, gene, phenotype and pathology data for glioma patients	50 patients/year
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	Based on input provided by pilots

A.4 Data utility and identification

Table 11: Purpose and re-use of data (data utility) and identification.

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
DS1.1	Cardiac ultrasound video recordings	Development and evaluation of AI-assisted cardiology diagnosis tool	Partially identifiable since it will be interlinked with a specific patient. We could interlink extracted features	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
			and medical images/videos.	pending publication or patent). Request must be issued by each individual partner to AHEPA.
DS1.2	Capsule endoscopy video recordings	Development and evaluation of AI-assisted gastroenterology diagnosis tool	Partially identifiable since it will be interlinked with a specific patient. We could interlink extracted features and medical images/videos.	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Request must be issued by each individual partner to AHEPA.
DS1.3	Cardiotocography variables and results, biometric data, medical history data	Development and evaluation of AI-assisted diagnosis tool	TBD, possibly anonymous data, but interconnected with hospital data in case of needed patient identification	It is not available/accessible outside pilot #1 AUTH.
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	Development and evaluation of AI-assisted diagnosis tool	TBD, possibly anonymous data, but interconnected with hospital data in case of needed patient identification	It is not available/accessible outside pilot #1 AHEPA/AUTH.
DS2.1	Data related to hours spent by specialists.	Create patient's schedule and assess patient's satisfaction	Anonymized	None
DS2.2	Retrospective patient schedule data and	Develop AI model and digital twin.	Anonymized	Under the conditions of the Data Processing Agreement

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
	precondition of the treatment			between the partners involved in the pilot #2
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	Develop AI model and digital twin Validation.	Fully identifiable	None
DS2.4	PROMs/PREMs	Validation (chatbot)	Anonymized	PROM and PREM questionnaire no restriction, however TBD how chatbot data will be shared with partners
DS2.5	Prospective patient clinical data	Validation	Fully identifiable, should not be shared.	Private patient information is not available/accessible outside pilot #2.
DS2.6	Patient personal data (address, preferences, …)	Validation	Fully identifiable, should not be shared.	Private patient information is not available/accessible outside pilot #2.
DS2.7	Retrospective EHR	Develop AI model and digital twin	Anonymized	Under the conditions of the Data Processing Agreement between the partners involved in pilot #2.

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction	Development algorithm for Optimization Scheduler.	Fully identifiable, should not be shared.	Private patient information is not available/accessible outside pilot #2. What we can share in extracted is features + pattern identification outcome since this is anonymized and will be also exploited in publications
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption logs, Human presence/access control, Devices' activation and connected loads	Monitor of the environment	Fully identifiable since it will be interlinked with a specific location	Data belongs to the owner of the location where the home automation devices are installed. The owner defines access to the data
DS3.2	IMU data captured by iPrognosis smartphone application	Remote monitoring of people with Parkinson's disease (PwP) via the iPrognosis application	Pseudo-anonymised	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Data access roles will be granted.
DS3.3	Voice-related time and spectral	Remote monitoring of PwP via the	Pseudo-anonymised	Access to pseudo-anonymised data is allowed with no

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
	features captured by iPrognosis smartphone application	iPrognosis application		restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Data access roles will be granted.
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	Remote monitoring of PwP via the iPrognosis application	Pseudo-anonymised	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Data access roles will be granted.
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	Remote assessment of motor status of PwP via the iPrognosis iMAT application	Pseudo-anonymised	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Data access roles will be granted.
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	Develop AI intervention model and Validation	Pseudo-anonymised	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Data access roles will be granted.

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
DS4.xx				
DS5.1	Audio-Visual recordings of patients	Support for Spoken Language Interaction, Classification of psychological distress (symptoms of depression) and action recognition	Fully identifiable, should not be shared. We can think of sharing extracted features	It is not available/accessible outside pilot #5 UM/UKCM. Special DTA is signed with UKCM as data owner and UM as data processor prior to the trial. ***What we can share is extracted features + classification outcome since this is anonymized and will be also exploited in publications***
DS5.2	Biometric data (e.g., blood pressure and heart rate)	Impact of PGHD on CCDS Efficiency and improved management of clinical parameters	Partially identifiable since it will be interlinked with a specific patient. We could interlink extracted features and biometric data as a data set.	It is not available/accessible outside pilot #5 UM/UKCM. Special DTA with UKCM as data owner and UM as data processor will be prepared and signed prior to the trial. **We will consider, however, creating a dataset correlating interaction features and monitored biomarkers with moods/emotions/psychological distress**

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
DS5.3	Retrospective electronic health records	Impact of CPOE on clinical routine, CCDS Efficiency and improved management of clinical parameters	Fully identifiable since it will be interlinked with a specific patient. Cannot be shared.	It is not available/accessible outside pilot #5 UM/UKCM. DTA with UKCM as data owner and UM as data processor will be prepared and signed prior to the trial.
DS5.4	PREMs related to clinical staff (depends on T1.4)	Impact of Social robotics system on various aspects of clinical workflow, including staff satisfaction and workload	Anonymized, however a code will be provided for comparison prior, during and after intervention	Statistical data and cohorts available to be shared with the consortium
DS5.5	PREMs related patients (e.g., PAM, SUS-SI/TAM, UEQ) and PROs	Impact of Social robotics system on quality of care	Anonymized, however a code will be provided for comparison prior, during and after intervention	Statistical data and cohorts available to be shared with the consortium
DS5.6	Datasets for facial expression and emotion recognition	Development of sensing AI to collect and classify symptoms of depression from facial expressions, speech and text.	Public dataset, please consider the individual licenses and restrictions	The data is openly available, request must be issued by each individual partner to the specific owner of the data set
DS5.7	Datasets for ASR and TTS in Slovenian	Support for Spoken Language Interaction	Proprietary dataset owned by UM. It is background we bring into the project.	The data is not publicly open. Access can be granted on individual basis (bilateral agreements which

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
				may include charges)
DS5.8	Datasets for ASR and TTS in French	Support for Spoken Language Interaction	Proprietary dataset owned by UM and public datasets. It is background we bring into the project.	Access to public datasets must be managed by individual partners, access to UM's closed datasets can be granted on individual basis (bilateral agreements which may include charges)
DS5.9	EVA Corpus, data set of conversational expression	Support for Spoken Language Interaction	Proprietary dataset owned by UM. It is background we bring into the project. It is already publicly available.	Access via CLARIN.SI repository
DS5.10	Video recordings of third persons	TBD	TBD	TBD
DS5.11	Patient's behavioural information	TBD	TBD	TBD
DS5.12	Patient's facial information	TBD	TBD	TBD
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data.	Support for identification of abnormal pattern recognition.	Fully identifiable, should not be shared.	Private patient information is not available/accessible outside pilot #6. What we can share in extracted is features + pattern identification outcome since this is anonymized and will be also

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
				exploited in publications
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Retrospective dataset starting from 11/2020 used to develop an AI prototype to alleviate the administrative burden in the interventional suite by an automatic procedure tracking.	Pseudo anonymization. Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)	Data will be not available/accessible outside pilot #7
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Retrospective dataset starting from 11/2020 used to develop an AI prototype to alleviate the administrative burden in the interventional suite by an automatic procedure tracking.	Pseudo anonymization. Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)	Data will be not available/accessible outside pilot #7
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	Retrospective dataset starting from 11/2020 used to develop an AI prototype to ensure an automatic logging a smart reporting of both imaging and patient X-ray dosimetry and a	Pseudo anonymization. Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)	Data will be not available/accessible outside pilot #7

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
		facilitated coronary angiogram interpretation by calculation of severity scores		
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	Retrospective dataset starting from 11/2020 used to develop an AI prototype to ensure a clinical decision support.	Pseudo anonymization. Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)	Data will be not available/accessible outside pilot #7
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	Retrospective dataset starting from 11/2020 used to develop an AI prototype to ensure a clinical decision support.	Pseudo anonymization. Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)	Data will be not available/accessible outside pilot #7
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive	Retrospective dataset starting from 11/2020 used to develop an AI prototype to ensure a	Pseudo anonymization. Data will be anonymized before sending to third party (Philips) but	Data will be not available/accessible outside pilot #7

Code	Type of data to be collected / name of the dataset	Purpose and re-use of the data (data utility)	Collected information will be identifiable?	How will the data be made accessible to other partners in consortium? Are there any restrictions?
	coronary angiogram and/or a coronary intervention	clinical decision support.	the encrypting key will be stored by the owner of the data (UZB, VUB)	
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	Prospective validation of AI prototype to ensure automatic procedure tracking, an automatic logging, a smart reporting and a help to clinical decision support.	Non anonymized data will be treated on site (UZB, VUB) by AI smartcathlab Philips prototype Data required to be sent for an extern analysis by a third party will be anonymized but the encrypting key will be stored by the owner of the data (UZB, VUB)	Data will be not available/accessible outside pilot #7
DS8.1	Image, gene, phenotype and pathology data for glioma patients	Dataset for highlighting relationships between different data types and identifying tumour type	The data will be identifiable and will not be shared as is, but procedures to make them (partially) available will be pursued	The data is restricted, we will pursue ways to make the data (partially) available to the consortium via pseudo anonymization; work is ongoing with UZ Brussel.
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	To identify the efficiency in terms of cost-effectiveness of the new AI technologies vs. The previous state of art of all pilot data	Patient level data will be anonymized, only the intervention will be known to the statistician/health economist.	The primary data is property of each pilot, and the results of the economic/PROM/PREM analysis will be available to all partner. Coding of the economic part is property of PhE.

A.5 Data findability, metadata

Table 12: Making data findable, metadata

Code	Type of data to be collected / name of the dataset	Data identification and versioning	Naming conventions	Search keywords for re-use	Metadata
DS1.1	Cardiac ultrasound video recordings	TBD	TBD	TBD	TBD
DS1.2	Capsule endoscopy video recordings	TBD	TBD	TBD	TBD
DS1.3	Cardiotocography variables and results, biometric data, medical history data	TBD	TBD	TBD	TBD
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	TBD	TBD	TBD	TBD
DS2.1	Data related to hours spent by specialists.	By date, by system (legacy vs AI). TBC whether this can be published or not.	TBD	Radiotherapy, throughput, scheduling, satisfaction	TBD
DS2.2	Retrospective patient schedule data and precondition of the treatment	All schedule changes are versioned. Those data should not be persisted/shared beyond the project's scope.	TBD	Not applicable	Not applicable

Code	Type of data to be collected / name of the dataset	Data identification and versioning	Naming conventions	Search keywords for re-use	Metadata
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	By site	TBD	Same as DS2.1	TBD
DS2.4	PROMs/PREMs	Those data should not be persisted/shared beyond the project's scope.	TBD	Not applicable	Not applicable
DS2.5	Prospective patient clinical data	Internal use only	TBD	Not applicable	Not applicable
DS2.6	Patient personal data (address, preferences, …)	Internal use only	TBD	Not applicable	Not applicable
DS2.7	Retrospective EHR	Those data should not be persisted/shared beyond the project's scope.	TBD	Not applicable	Not applicable
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction	TBD	TBD	TBD	TBD
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption	TBD	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Data identification and versioning	Naming conventions	Search keywords for re-use	Metadata
	logs, Human presence/access control, Devices' activation and connected loads				
DS3.2	IMU data captured by iPrognosis smartphone application	TBD	TBD	TBD	TBD
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	TBD	TBD	TBD	TBD
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	TBD	TBD	TBD	TBD
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	TBD	TBD	TBD	TBD
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	TBD	TBD	TBD	TBD
DS4.xx	TBD				
DS5.1	Audio-Visual recordings of individuals	Patient, and Department, ID (internal hospital ID), Date-Time, Unit ID	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Data identification and versioning	Naming conventions	Search keywords for re-use	Metadata
DS5.2	Biometric data (e.g. blood pressure and heart rate)	Patient, and Department, ID (internal hospital ID), Date-Time, Monitor ID	TBD	TBD	TBD
DS5.3	Retrospective electronic health records	N/A	N/A	N/A	N/A
DS5.4	PREMs related to clinical staff (depends on T1.4)	Staff-ID, Department ID, Date-Time	TBD	TBD	TBD
DS5.5	PREMs related patients (e.g. PAM, SUS-SI/TAM, UEQ) and PROs	Patient, and Department, ID (internal hospital ID), Date-Time	TBD	TBD	TBD
DS5.6	Datasets for facial expression and emotion recognition	N/A	N/A	N/A	N/A
DS5.7	Datasets for ASR and TTS in Slovenian	N/A	N/A	N/A	N/A
DS5.8	Datasets for ASR and TTS in French	N/A	N/A	N/A	N/A
DS5.9	EVA Corpus, data set of conversational expression	N/A	N/A	N/A	N/A
DS5.10	Video recording of third persons	By robot ID and date	TBD	TBD	TBD
DS5.11	Patient's behavioural information	By patient name or ID, robot ID and date	TBD	TBD	TBD
DS5.12	Patient's facial information	By patient name or ID	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Data identification and versioning	Naming conventions	Search keywords for re-use	Metadata
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data.	Age, Body Height, General status. Heartbeat. Oxygen, Blood pressure, Body temperature, Urine, Glucose, Body weight, Stool.	Heartbeat: X' Oxygen: % Blood pressure: mmHg Body temperature: ºC Urine: chromatic scale Glucose: mg-DL Body weight: Kg Stool: Number of times per day.	TBD	TBD
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD
DS7.4	Coronary physiology data	TBD	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Data identification and versioning	Naming conventions	Search keywords for re-use	Metadata
	of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB				
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	TBD	TBD	TBD	TBD
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	TBD	TBD	TBD	TBD
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	TBD	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Data identification and versioning	Naming conventions	Search keywords for re-use	Metadata
DS8.1	Image, gene, phenotype and pathology data for glioma patients	Unique IDs via patient numbers, versioning for local systems exists, TBD for central one	TBD	TBD	TBD

A.6 Data open accessibility

Table 13: Making data open accessible.

Code	Type of data to be collected / name of the dataset	Dataset made available openly?	Which open repository (or other available medium)?	Licenses, access identification, restrictions
DS1.1	Cardiac ultrasound video recordings	YES (unless otherwise decided due to IPRs)	Zenodo	TBD
DS1.2	Capsule endoscopy video recordings	YES (unless otherwise decided due to IPRs)	Zenodo	TBD
DS1.3	Cardiotocography variables and results, biometric data, medical history data	TBD	TBD	TBD
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	TBD	TBD	TBD
DS2.1	Data related to hours spent by specialists.	Yes	TBD	TBD

Code	Type of data to be collected / name of the dataset	Dataset made available openly?	Which open repository (or other available medium)?	Licenses, access identification, restrictions
DS2.2	Retrospective patient schedule data and precondition of the treatment	All schedule changes are versioned. Those data should not be persisted/shared beyond the project's scope.	TBD	Not applicable
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	Same as DS2.1		
DS2.4	PROMs/PREMs	Those data should not be persisted/shared beyond the project's scope.	Not applicable	Not applicable
DS2.5	Prospective patient clinical data	Internal use only	Not applicable	Not applicable
DS2.6	Patient personal data (address, preferences, …)	Internal use only	Not applicable	Not applicable
DS2.7	Retrospective EHR	Those data should not be persisted/shared beyond the project's scope.	Not applicable	Not applicable
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction	NO	N/A	N/A

Code	Type of data to be collected / name of the dataset	Dataset made available openly?	Which open repository (or other available medium)?	Licenses, access identification, restrictions
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption logs, Human presence/access control, Devices activation and connected loads	Data belongs to the owner of the location where the home automation devices are installed. The owner defines access to the data.	None	The owner defines access to the data.
DS3.2	IMU data captured by iPrognosis smartphone application	YES (unless otherwise decided due to IPRs)	Zenodo	TBD
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	YES (unless otherwise decided due to IPRs)	Zenodo	TBD
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	YES (unless otherwise decided due to IPRs)	Zenodo	TBD
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	YES (unless otherwise decided due to IPRs)	Zenodo	TBD
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	YES (but considering the limitations related to the IPR/Consortium Agreement)	TBD	TBD

Code	Type of data to be collected / name of the dataset	Dataset made available openly?	Which open repository (or other available medium)?	Licenses, access identification, restrictions
DS4.xx	TBD			
DS5.1	Audio-Visual recordings of patients	NO	N/A	N/A
DS5.2	Biometric data (e.g., blood pressure and heart rate)	TBD	TBD	TBD
DS5.3	Retrospective electronic health records	TBD	TBD	TBD
DS5.4	PREMs related to clinical staff (depends on T1.4)	YES (but considering anonymization)	TBD	TBD
DS5.5	PREMs related patients (e.g., PAM, SUS-SI/TAM, UEQ) and PROs	YES (but considering anonymization)	TBD	TBD
DS5.6	Datasets for facial expression and emotion recognition	YES	e.g., jaffe, FER-2013, MMI, Cohn-Kanade, RaFD, FERG, EMOTIC, Affect data, SemEval-2017 Task 4, DailyDialog, The MPLab GENKI Database, FFECTIVA-MIT Facial Expression Dataset (AM-FED), Grounded	Licenses are granted on individual requests, not handled by UM but the owners of data sets

Code	Type of data to be collected / name of the dataset	Dataset made available openly?	Which open repository (or other available medium)?	Licenses, access identification, restrictions
			Emotions, Reldi, etc.	
DS5.7	Datasets for ASR and TTS in Slovenian	NO		Access can be granted on individual basis (bilateral agreements which may include charges)
DS5.8	Datasets for ASR and TTS in French	NO		Access can be granted on individual basis (bilateral agreements which may include charges)
DS5.9	EVA Corpus, data set of conversational expression	YES	Clarin.SI	CC-BY 4.0 License
DS5.10	Video recordings of third persons	NO	N/A	N/A
DS5.11	Patient's behavioural information	NO	N/A	N/A
DS5.12	Patient's facial information	NO	N/A	N/A
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data.	NO	N/A	N/A
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD
DS7.2	Clinical data of patients	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Dataset made available openly?	Which open repository (or other available medium)?	Licenses, access identification, restrictions
	scheduled for a coronary angiogram/coron ary intervention @ UZB			
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coron ary intervention @ UZB	TBD	TBD	TBD
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	TBD	TBD	TBD
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	TBD	TBD	TBD
DS7.6	Coronary CT data including FFRCT computation of	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Dataset made available openly?	Which open repository (or other available medium)?	Licenses, access identification, restrictions
	patient referred for an invasive coronary angiogram and/or a coronary intervention			
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	TBD	TBD	TBD
DS8.1	Image, gene, phenotype and pathology data for glioma patients	No	Not relevant	TBD
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	Applicable for pilot specifics	Applicable for pilot specifics	Applicable for pilot specifics

A.7 Data interoperability

Table 14: Data interoperability.

Code	Type of data to be collected / name of the dataset	Metadata vocabularies, ontologies, standards and methodologies for data interoperability
DS1.1	Cardiac ultrasound video recordings	HL7
DS1.2	Capsule endoscopy video recordings	HL7
DS1.3	Cardiotocography variables and results, biometric data, medical history data	TBD

Code	Type of data to be collected / name of the dataset	Metadata vocabularies, ontologies, standards and methodologies for data interoperability
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	TBD
DS2.1	Data related to hours spent by specialists.	TBD
DS2.2	Retrospective patient schedule data and precondition of the treatment	TBD
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	TBD
DS2.4	PROMs/PREMs	TBD
DS2.5	Prospective patient clinical data	TBD
DS2.6	Patient personal data (address, preferences, …)	TBD
DS2.7	Retrospective EHR	TBD
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction	N/A
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption logs, Human presence/access control, Devices' activation and connected loads	N/A
DS3.2	IMU data captured by iPrognosis smartphone application	HL7
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	HL7
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	HL7
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	HL7

Code	Type of data to be collected / name of the dataset	Metadata vocabularies, ontologies, standards and methodologies for data interoperability
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	TBD
DS4.xx	TBD	
DS5.1	Audio-Visual recordings of patients	FHIR v4
DS5.2	Biometric data (e.g., blood pressure and heart rate)	FHIR v4
DS5.3	Retrospective electronic health records	N/A
DS5.4	PREMs related to clinical staff (depends on T1.4)	TBD
DS5.5	PREMs related patients (e.g., PAM, SUS-SI/TAM, UEQ) and PROs	FHIR v4
DS5.6	Datasets for facial expression and emotion recognition	N/A
DS5.7	Datasets for ASR and TTS in Slovenian	
DS5.8	Datasets for ASR and TTS in French	N/A
DS5.9	EVA Corpus, data set of conversational expression	N/A
DS5.10	Video recordings of third persons	N/A
DS5.11	Patient's behavioural information	FHIR v4
DS5.12	Patient's facial information	FHIR v4
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data.	N/A
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	TBD

Code	Type of data to be collected / name of the dataset	Metadata vocabularies, ontologies, standards and methodologies for data interoperability
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	TBD
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	TBD
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	TBD
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	TBD
DS8.1	Image, gene, phenotype and pathology data for glioma patients	TBD
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	Applicable for pilot specifics

A.8 Data licensing, availability and usability

Table 15: Data licensing, availability and usability by third parties.

Code	Type of data to be collected / name of the dataset	Availability and usability of data by third parties. Licensing
DS1.1	Cardiac ultrasound video recordings	TBD
DS1.2	Capsule endoscopy video recordings	TBD
DS1.3	Cardiotocography variables and results, biometric data, medical history data	TBD
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	TBD

Code	Type of data to be collected / name of the dataset	Availability and usability of data by third parties. Licensing
DS2.1	Data related to hours spent by specialists.	Consortium agreement
DS2.2	Retrospective patient schedule data and precondition of the treatment	Joint controller processing agreement
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	Consortium agreement
DS2.4	PROMs/PREMs	Internal use only
DS2.5	Prospective patient clinical data	Internal use only
DS2.6	Patient personal data (address, preferences, …)	Internal use only
DS2.7	Retrospective EHR	Joint controller processing agreement
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction.	N/A
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption logs, Human presence/access control, Devices' activation and connected loads	TBD
DS3.2	IMU data captured by iPrognosis smartphone application	TBD
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	TBD
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	TBD
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	TBD
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	Consortium agreement
DS4.xx	TBD	
DS5.1	Audio-Visual recordings of patients	N/A
DS5.2	Biometric data (e.g., blood pressure and heart rate)	TBD

Code	Type of data to be collected / name of the dataset	Availability and usability of data by third parties. Licensing
DS5.3	Retrospective electronic health records	N/A
DS5.4	PREMs related to clinical staff **(**depends on T1.4**)**	TBD
DS5.5	PREMs related patients (e.g., PAM, SUS-SI/TAM, UEQ) and PROs	TBD
DS5.6	Datasets for facial expression and emotion recognition	Licenses are granted on individual requests, not handled by UM but the owners of data sets
DS5.7	Datasets for ASR and TTS in Slovenian	Access can be granted on individual basis (bilateral agreements which may include charges)
DS5.8	Datasets for ASR and TTS in French	Access can be granted on individual basis (bilateral agreements which may include charges)
DS5.9	EVA Corpus, data set of conversational expression	CC-BY 4.0 License
DS5.10	Video recordings of third persons	N/A
DS5.11	Patient's behavioural information	N/A
DS5.12	Patient's facial information	N/A
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data.	N/A
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	TBD
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual	TBD

Code	Type of data to be collected / name of the dataset	Availability and usability of data by third parties. Licensing
	or a motorized wire pullback and performed before and after a coronary intervention @ UZB	
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	TBD
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	TBD
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	TBD
DS8.1	Image, gene, phenotype and pathology data for glioma patients	TBD
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	Applicable for pilot specifics

A.9 Data quality assurance

Table 16: Data quality assurance

Pilot nr	Task nr	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
1	5.2	TBD	TBD	AUTH; person(s) TBD	Peer-review
2	3.2	TBD	Verification and validation plan	CHU de Liège, Patrick Duflot	System test, Unit test, Manual test
2	5.2	TBD	Monitoring Plan ensuring that the collected data	CHU de Liège, Patrick Duflot	TBD in study protocol

Pilot nr	Task nr	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
			are the right ones and completed.		
2	5.3	As defined by PhE	As defined by PhE	As defined by PhE	As defined by PhE
3	3.3	TBD	TBD	TBD	TBD
4	3.5	TBD	TBD	INTRAS (person(s) TBD)	TBD
5	3.5	TBD	TBD	ITCL	TBD
6	3.5	Protocols shared by PhE	Protocols shared by PhE	ITCL	TBD
7	3.6	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
7	3.6	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
7	3.6	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD

Pilot nr	Task nr	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
		intervention @ UZB			
7	3.6	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
7	3.6	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention@ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
7	3.6	Coronary CT data including FFRCT computation of patient referred for an	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD

Pilot nr	Task nr	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
		invasive coronary angiogram and/or a coronary intervention@ UZB			
7	3.6	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT of data.	Prospective study protocol (comparative observational study) is under preparation	Primary investigator UZB, VUB (TBD)	Comparative analysis of different key performance indicators and patient reported outcome and experience measures (PREMS and PROMS) evolution before (control period) and after implantation of AI smartcathlab prototype (study period)
8	N/A	Image, gene, phenotype and pathology data for glioma patients	To be determined in collaboration with UZ Brussels ICT	VUB; to be determined	Local testing, feedback from specialists, eventual peer review
PhE	5.3	Based on Quality control Standard Operating Procedure	Protocol & CRF has been prepared by PhE and will be shared to	PhE project team / Eugena Stamuli, Declan O'	TBD

Pilot nr	Task nr	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
		(SOP) followed by PhE for all projects/deliverables	all pilots in order to follow the same	Byrne, Magda Chatzikou	

A.10 Data cleansing, transforming and analysing

Table 17: Data cleansing, transforming and analysing.

Code	Type of data to be collected / name of the dataset	Type of data cleaning needed (e.g., correct data types, remove duplicates, add missing info…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g., normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
DS1.1	Cardiac ultrasound video recordings	TBD	TBD	TBD	TBD	TBD	TBD
DS1.2	Capsule endoscopy video recordings	TBD	TBD	TBD	TBD	TBD	TBD
DS1.3	Cardiotocography variables and results, biometric data, medical history data	TBD	TBD	TBD	TBD	TBD	TBD
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	TBD	TBD	TBD	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Type of data cleaning needed (e.g., correct data types, remove duplicates, add missing info…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g., normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
DS2.1	Data related to hours spent by specialists.	TBD	TBD	TBD	TBD	TBD	TBD
DS2.2	Retrospective patient schedule data and precondition of the treatment	TBD	TBD	TBD	TBD	TBD	TBD
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	TBD	TBD	TBD	TBD	TBD	TBD
DS2.4	PROMs/PREMs	TBD	TBD	TBD	TBD	TBD	TBD
DS2.5	Prospective patient clinical data	TBD	TBD	TBD	TBD	TBD	TBD
DS2.6	Patient personal data (address, preferences, …)	TBD	TBD	TBD	TBD	TBD	TBD
DS2.7	Retrospective EHR	TBD	TBD	TBD	TBD	TBD	TBD
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction	N/A	TBD	TBD	TBD	TBD	TBD
DS3.1	Smart home data: Consumption/ production of instantaneous electricity and consumption logs, Human presence/acces	Outlier detection, inference on missing data	VIMAR	Normalization	TBD	TBD	TBD

Code	Type of data to be collected / name of the dataset	Type of data cleaning needed (e.g., correct data types, remove duplicates, add missing info…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g., normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
	s control, Devices' activation and connected loads						
DS3.2	IMU data captured by iPrognosis smartphone application	TBD	TBD	TBD	TBD	TBD	TBD
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	TBD	TBD	TBD	TBD	TBD	TBD
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	TBD	TBD	TBD	TBD	TBD	TBD
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	TBD	TBD	TBD	TBD	TBD	TBD
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	TBD	TBD	TBD	TBD	TBD	TBD
DS4.xx	TBD						

Code	Type of data to be collected / name of the dataset	Type of data cleaning needed (e.g., correct data types, remove duplicates, add missing info…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g., normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
DS5.1	Audio-Visual recordings of patients	N/A	N/A	N/A	N/A	N/A	N/A
DS5.2	Biometric data (e.g., blood pressure and heart rate)	Data cleaning will be performed by data controller (UKCM)	TBD	TBD	TBD	TBD	TBD
DS5.3	Retrospective electronic health records	N/A	N/A	N/A	N/A	N/A	N/A
DS5.4	PREMs related to clinical staff (depends on T1.4)	Data cleaning will be performed by data controller (UKCM)	TBD	TBD	TBD	TBD	TBD
DS5.5	PREMs related patients (PAM, SUS-SI/TAM, UEQ)	Data cleaning will be performed by data controller (UKCM)	TBD	TBD	TBD	TBD	TBD
DS5.6	Datasets for facial expression and emotion recognition	N/A	N/A	N/A	N/A	N/A	N/A
DS5.7	Datasets for ASR and TTS in Slovenian	N/A	N/A	N/A	N/A	N/A	N/A
DS5.8	Datasets for ASR and TTS in French	N/A	N/A	N/A	N/A	N/A	N/A
DS5.9	EVA Corpus, data set of conversational expression	N/A	N/A	N/A	N/A	N/A	N/A

Code	Type of data to be collected / name of the dataset	Type of data cleaning needed (e.g., correct data types, remove duplicates, add missing info…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g., normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
DS5.10	Video recordings of third persons	N/A	N/A	N/A	N/A	N/A	N/A
DS5.11	Patient's behavioural information	TBD	TBD	TBD	TBD	TBD	TBD
DS5.12	Patient's facial information	TBD	TBD	TBD	TBD	TBD	TBD
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data.	TBD	TBD	TBD	TBD	TBD	TBD
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Excel	Philips Image Guided Therapy Systems, Netherlands	TBD
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Excel	Philips Image Guided Therapy Systems, Netherlands	TBD
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Excel	Philips Image Guided Therapy Systems, Netherlands	TBD

Code	Type of data to be collected / name of the dataset	Type of data cleaning needed (e.g., correct data types, remove duplicates, add missing info…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g., normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low physiology curves quality)	Coroventis software and Virtual stenting algorithm (VSA) for interoperation of RFR/FFR pullback created by JF Argacha and Jean Decamp (I-depot number 123060 Date 16-04-2020)	Philips Image Guided Therapy Systems, Netherlands And JF Argacha, cardiology department, UZB, VUB, Brussel	TBD
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Optis and volcano software	Philips Image Guided Therapy Systems, Netherlands	TBD
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Philips and Heartflow software	Philips Image Guided Therapy Systems, Netherlands	TBD
DS7.7	Full prospective	Not applicable	Not applicable	Not applicable	Not applicable	Philips Image Guided	TBD

Code	Type of data to be collected / name of the dataset	Type of data cleaning needed (e.g., correct data types, remove duplicates, add missing info…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g., normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
	UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	(prospective inclusion)	(prospective inclusion)	(prospective inclusion)	(prospective inclusion)	Therapy Systems, Netherlands And Cardiology department UZB, VUB	
DS8.1	Image, gene, phenotype and pathology data for glioma patients	Connecting sample information per patient across databases, machine learning on data	To be hired	Integration, neural networks	Python, pytorch, sklearn, XNAT framework	VUB	To be determined
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	Data cleaning will be performed by each data owner (pilot). PhE will perform the analysis on clean datasets	TBD by each pilot.	Economic & PRO/PREM analysis (economic evaluation, cost consequence analysis, cost-utility analysis, patients' quality of life measurement, bootstrapping, regression, etc	Stata, Excel, maybe SAS if necessary	PhE	Following ISPOR guidelines for economic evaluation and Dolan's publication on EQ-5D analysis

A.11 Ethical review

Table 18: Ethical review.

Code	Type of data to be collected / name of the dataset	Type of ethical review needed
DS1.1	Cardiac ultrasound video recordings	Clinical protocol, Ethical approval by Institutional Research Ethics Board
DS1.2	Capsule endoscopy video recordings	Clinical protocol, Ethical approval by Institutional Research Ethics Board
DS1.3	Cardiotocography variables and results, biometric data, medical history data	Clinical protocol and ethical approval from the hospital administration
DS1.4	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	Clinical protocol and ethical approval from the hospital administration
DS2.1	Data related to hours spent by specialists.	N/A
DS2.2	Retrospective patient schedule data and precondition of the treatment	DPO approval
DS2.3	Data linked to radiotherapy machines (tumours treatment indication, maintenance, building location)	N/A
DS2.4	PROMs/PREMs	Ethical approval
DS2.5	Prospective patient clinical data	Ethical approval
DS2.6	Patient personal data (address, preferences, …)	Ethical approval
DS2.7	Retrospective EHR	DPO approval
DS2.8	Data from radiotherapy services, infrastructure. Patient's satisfaction	N/A
DS3.1	Smart home data: Consumption/production of instantaneous electricity and consumption logs, Human presence/access control, Devices' activation and connected loads	Clinical protocol, Ethical approval by national review board
DS3.2	IMU data captured by iPrognosis smartphone application	Clinical protocol, Ethical approval by Research Ethics Committee
DS3.3	Voice-related time and spectral features captured by iPrognosis smartphone application	Clinical protocol, Ethical approval by Research Ethics Committee
DS3.4	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	Clinical protocol, Ethical approval by Research Ethics Committee

Code	Type of data to be collected / name of the dataset	Type of ethical review needed
DS3.5	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	Clinical protocol, Ethical approval by Research Ethics Committee
DS3.6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	Clinical protocol, Ethical approval by Research Ethics Committee
DS4.xx	TBD	
DS5.1	Audio-Visual recordings of patients	N/A
DS5.2	Biometric data (e.g., blood pressure and heart rate)	Clinical protocol, Ethical approval by national review board
DS5.3	Retrospective electronic health records	Clinical protocol, Ethical approval by national review board
DS5.4	PREMs related to clinical staff (depends on T1.4)	Clinical protocol, Ethical approval by national review board
DS5.5	PREMs related patients (e.g., PAM, SUS-SI/TAM, UEQ) and PROs	Clinical protocol, Ethical approval by national review board
DS5.6	Datasets for facial expression and emotion recognition	N/A
DS5.7	Datasets for ASR and TTS in Slovenian	N/A
DS5.8	Datasets for ASR and TTS in French	N/A
DS5.9	EVA Corpus, data set of conversational expression	N/A
DS5.10	Video recordings of third persons	N/A
DS5.11	Patient's behavioural information	N/A
DS5.12	Patient's facial information	N/A
DS6.1	Patient's physical, medical and mental status. Vital signs with sensors data	N/A
DS7.1	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Clinical protocol (patient registry). Ethical approval by hospital review board will be asked.

Code	Type of data to be collected / name of the dataset	Type of ethical review needed
DS7.2	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Clinical protocol (patient registry). Ethical approval by hospital review board will be asked.
DS7.3	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	Clinical protocol (patient registry). Ethical approval by hospital review board will be asked.
DS7.4	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	Clinical protocol (patient registry). Ethical approval by hospital review board will be asked.
DS7.5	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	Clinical protocol (patient registry). Ethical approval by hospital review board will be asked.
DS7.6	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	N/A
DS7.7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	Clinical protocol (prospective comparative study). Ethical approval by hospital review board will be asked.
DS8.1	Image, gene, phenotype and pathology data for glioma patients	Already performed and approved.
DS.O.1	All pilot data with Common KPIs (Economic & PROMs/PREMs)	The economic & PROMs/PREMs data will be included in the protocol.

Appendix B The Data Management Survey

a. Please describe which projects' partners will be collecting data in their WPs, and link to the WP tasks. Please be specific for each type of data collection.

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
5.2	5	Audio-Visual recordings of individuals	Support for Spoken Language Interaction, Classification of psychological distress (symptoms of depression) and action recognition	UM	UKCM, ITCL	At least 1 recording per day per patient	Clinical protocol, Ethical approval by national review board	Fully identifiable, should not be shared. We can think of sharing extracted features
5.2	5	Biometric data (e.g., blood pressure and heart rate)	Impact of PGHD on CCDS Efficiency and improved management of clinical parameters	UKCM	UM	N/A	Clinical protocol, Ethical approval by national review board	Partially identifiable since it will be interlinked with a specific patient. We could interlink extracted features

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
								and biometric data as a data set.
5.2	5	Retrospective electronic health records	Impact of CPOE on clinical routine, CCDS Efficiency and improved management of clinical parameters	UKCM	UM	Up to 100MB per patient	Clinical protocol, Ethical approval by national review board	Fully identifiable since it will be interlinked with a specific patient. Cannot be shared.
5.2	5	PREMs related to clinical staff **(**depends on T1.4**)**	Impact of Social robotics system on various aspects of clinical workflow, including staff satisfaction and workload	UKCM	UM	N/A	Clinical protocol, Ethical approval by national review board	Anonymized, however a code will be provided for comparison prior, during and after intervention
5.2	5	PREMs related patients (PAM, SUS-SI/TAM, UEQ)	Impact of Social robotics system	UKCM	UM	N/A	Clinical protocol, Ethical	Anonymized, however a code will be provided for

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
			on quality of care				approval by national review board	comparison prior, during and after intervention
3.5	5	Datasets for facial expression and emotion recognition	Development of sensing AI to collect and classify symptoms of depression from facial expressions, speech and text.	UM	N/A		N/A	Public dataset, please consider the individual licenses and restrictions
3.5	5	Datasets for ASR and TTS in Slovenian	Support for Spoken Language Interaction	UM	N/A		N/A	Proprietary dataset owned by UM. It is background we bring into the project.
3.5	2	Datasets for ASR and TTS in French	Support for Spoken Language Interaction	UM				Proprietary dataset owned by UM and public datasets.

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
								It is background we bring into the project.
3.5	5	EVA Corpus, data set of conversational expression	Support for Spoken Language Interaction	UM	N/A		N/A	Proprietary dataset owned by UM. It is background we bring into the project. It is already publicly available.
5.2	1	Cardiac ultrasound video recordings	Development and evaluation of AI-assisted cardiology diagnosis tool	AHEPA	AUTH	TBD	Clinical protocol, Ethical approval by Institutional Research Ethics Board	Partially identifiable since it will be interlinked with a specific patient. We could interlink extracted features and medical images/videos.

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
5.2	1	Capsule endoscopy video recordings	Development and evaluation of AI-assisted gastroenterology diagnosis tool	AHEPA	AUTH	Recordings from 60 patients	Clinical protocol, Ethical approval by Institutional Research Ethics Board	Partially identifiable since it will be interlinked with a specific patient. We could interlink extracted features and medical images/videos.
5.2	3, 6	Inertial Measurement Unit (IMU) data captured by iPrognosis smartphone application	Remote monitoring of people with Parkinson's disease (PwP) via the iPrognosis application	AUTH	INTRAS	TBD	Clinical protocol, Ethical approval by Research Ethics Committee	Pseudo-anonymised
5.2	3, 6	Voice-related time and spectral features captured by iPrognosis smartphone application	Remote monitoring of PwP via the iPrognosis application	AUTH	INTRAS	TBD	Clinical protocol, Ethical approval by Research	Pseudo-anonymised

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
							Ethics Committee	
5.2	3, 6	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	Remote monitoring of PwP via the iPrognosis application	AUTH	INTRAS	TBD	Clinical protocol, Ethical approval by Research Ethics Committee	Pseudo-anonymised
5.2	3, 6	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	Remote assessment of motor status of PwP via the iPrognosis iMAT application	AUTH	INTRAS	TBD	Clinical protocol, Ethical approval by Research Ethics Committee	Pseudo-anonymised
5.2	6	Results of rehabilitation sessions with Gradior	Develop AI intervention	INTRAS	AUTH/ITCL	TBD	Clinical protocol, Ethical	Pseudo-anonymised

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
		and care plans (editor and patient data)	model and Validation				approval by Research Ethics Committee	
5.2	1	Cardiotocography variables and results, biometric data, medical history data	Development and evaluation of AI-assisted diagnosis tool	AUTH	-	TBD	Clinical protocol and ethical approval from the hospital administration	TBD, possibly anonymous data, but interconnected with hospital data in case of needed patient identification
5.2	1	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	Development and evaluation of AI-assisted diagnosis tool	AHEPA	AUTH	TBD	Clinical protocol and ethical approval from the hospital administration	TBD, possibly anonymous data, but interconnected with hospital data in case of needed patient identification
3.3	3	Smart home data: Consumption/production of instantaneous	Monitor of the environment	IRCCS	VIMAR	N/A	Clinical protocol, Ethical	Fully identifiable since it will be

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
		electricity and consumption logs, Human presence/access control, Devices activation and connected loads					approval by national review board	interlinked with a specific location
5.3	PhE	All pilot data with Common KPIs (Economic & PROMs/PREMs)	To identify the efficiency in terms of cost-effectiveness of the new AI technologies vs. The previous state of the art of all pilot data	All pilots	PhE	Based on input provided by pilots	The economic & PROMs/PREMs data will be included in the protocol which will be submitted by the pilots for ethical approval	Patient level data will be anonymized, only the intervention will be known to the statistician/health economist.
3.2	2	Data related to hours spent by specialists.	Create patient's schedule and assess patient's satisfaction	CHUL	ITCL/TMA	TBD	N/A	Anonymized

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
3.2	2	Retrospective patient schedule data and precondition of the treatment	Develop AI model and digital twin.	CHUL	ITCL/TMA	2500 patients/year	DPO approval	Anonymized
1.2	2	Data linked to radiotherapy machines (tumour treatment indication, maintenance, building location)	Develop AI model and digital twin Validation.	CHUL		Data of 5 machines	N/A	Fully identifiable
5.2	2	PROMs/PREMs	Validation (chatbot)	CHUL	UKCM	TBD	Ethical approval	Anonymized
5.2	2	Prospective patient clinical data	Validation	CHUL	ICTL/TMA	TBD	Ethical approval	Anonymized
5.2	2	Patient personal data (address, preference,)	Validation	CHUL	ICTL/TMA	TBD	Ethical approval	Anonymized

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
3.2 and 5.2	2	Retrospective EHR	Develop AI model and digital twin.	CHUL	ICTL/TMA	Subset to be identified if needed	DPO approval	Anonymized
3.5	6	Patient's physical, medical and mental status. Vital signs with sensors data.	Support for identification of abnormal pattern recognition.	ITCL	-	At least 1 recording per day per patient and per signal.	-	Fully identifiable, should not be shared.
3.2	2	Data from radiotherapy services, infrastructure. Patient's satisfaction	Development algorithm for Optimization Scheduler.	ITCL	TMA	TBD	-	Fully identifiable, should not be shared.
? Not sure what to put here	8	Image, gene, phenotype and pathology data for glioma patients	Dataset for highlighting relationships between different data types and identifying tumor type	VUB		50 patients/year	Already performed and approved	The data will be identifiable and will not be shared as is, but procedures to make them (partially) available will be pursued

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
3.6	7	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Retrospective dataset starting from 11/2020 used to develop an AI prototype to alleviate the administrative burden in the interventional suite by an automatic procedure tracking.	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)	Around 1100 coronary angiogram including 350 CA with PCI performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021	Clinical protocol (patient registry). Ethical approval by hospital review board will be asked.	Pseudoanonymization Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)
3.6	7	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Retrospective dataset starting from 11/2020 used to develop an AI prototype to alleviate the administrative burden in the	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)	Around 1100 coronary angiogram including 350 CA with PCI performed in cathlab 3	Clinical protocol (patient registry). Ethical approval by hospital	Pseudoanonymization Data will be anonymized before sending to third party (Philips) but

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
			interventional suite by an automatic procedure tracking.			and 4 between 01/11/2020 and 01/05/2021	review board will be asked.	the encrypting key will be stored by the owner of the data (UZB, VUB)
3.6	7	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	Retrospective dataset starting from 11/2020 used to develop an AI prototype to ensure an automatic logging a smart reporting of both imaging and patient X-ray dosimetry and a facilitated coronary angiogram interpretation	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)	Around 1100 coronary angiogram including 350 CA with PCI performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021	Clinical protocol (patient registry). Ethical approval by hospital review board will be asked.	Pseudoanonymization Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
			by calculation of severity scores					
3.6	7	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	Retrospective dataset starting from 11/2020 used to develop an AI prototype to ensure a clinical decision support.	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)	Around 150 single point and 20 motorized pullback of physiological evaluations by FFR/iFR/RFR performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021	Clinical protocol (patient registry). Ethical approval by hospital review board will be asked.	Pseudoanonymization Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)
3.6	7	Intravascular imaging data of patient evaluated by either OCT	Retrospective dataset starting from 11/2020	Cardiology department of	Philips Image Guided	Around 20 OCT and 10 IVUS	Clinical protocol	Pseudoanonymization

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
		or IVUS technique before and after a coronary intervention	used to develop an AI prototype to ensure an clinical decision support.	University Hospital Brussels, VUB	Therapy Systems (Philips)	performed in cathlab 3 and 4 between 01/11/2020 and 01/05/2021	(patient registry). Ethical approval by hospital review board will be asked.	Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)
3.6	7	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention	Retrospective dataset starting from 11/2020 used to develop an AI prototype to ensure a clinical decision support.	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)	Around 300 CTA performed in patient referred for an invasive angiogram performed in cathlab 3 and 4 between 01/11/2020		Pseudoanonymization Data will be anonymized before sending to third party (Philips) but the encrypting key will be stored by the owner of the data (UZB, VUB)

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
						and 01/05/2021		
3.6	7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	Prospective validation of AI prototype to ensure automatic procedure tracking, an automatic logging, a smart reporting and a help to clinical decision support.	Cardiology department of University Hospital Brussels, VUB	Philips Image Guided Therapy Systems (Philips)	Validation study of the AI prototype effects on key performance indicators (hospital and health care productivity). Prospective cohort of 100 patients.	Clinical protocol (prospective comparative study). Ethical approval by hospital review board will be asked.	Non anonymized data will be treated on site (UZB, VUB) by AI smartcathlab Philips prototype Data required to be sent for an extern analysis by a third party will be anonymized but the encrypting key will be stored by the owner of the data (UZB, VUB)
	4	TBD			91			

Task Nr.	Pilot Nr.	Type of data to be collected, and name of the dataset	Purpose and re-use of the data (data utility)	Responsible partner	Collaborating partners	Record size and expected data volume (n. of records)	Type of ethical review needed	Collected information will be identifiable?
3.5	5	Video recordings of the patient	Input for the action recognition algorithms	ITCL	N/A	TBD	-	Fully identifiable, should not be shared
3.5	5	Video recordings of third persons	Byproduct of the patient's recordings.	ITCL	N/A	TBD	-	Anonymous
3.5	5	Patient's behavioural information	Output of the action recognition algorithms. Support for the alert system.	ITCL	UM	TBD	-	Fully identifiable, should not be shared
3.5	5	Patient's facial information	Support for facial identification.	ITCL	N/A	At least 1 face image per patient	-	Fully identifiable, should not be shared

b. Please specify where the collected data will be stored for each task and type of data collection.

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
5.2	5	Audio-Visual recordings of patients in the trial	At the edge, dedicated infrastructure for the pilot	Open FHIR Server	Media Resource of FHIR Observation	It is not available/accessible outside pilot #5 UM/UKCM. Special DTA is signed with UKCM as data owner and UM as data processor prior to the trial. ***What we can share in extracted is features + classification outcome since this is anonymized and will be also exploited in publications***
5.2	5	Biometric data from patients in the trial (e.g. blood pressure and heart rate)	At the edge, dedicated infrastructure for the pilot	Open FHIR Server	FHIR Diagnostic Report, and Observation resource	It is not available/accessible outside pilot #5 UM/UKCM. Special DTA with UKCM as data owner and UM as data processor will be prepared and signed prior to the trial. **We will consider, however, creating a dataset correlating interaction features and monitored biomarkers with**

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
						moods/emotions/psychological distress
5.2	5	Retrospective electronic health records	At the edge, interface with existing IT system of the hospital	Proprietary Hl7 Compliant Medis Server	N/A	It is not available/accessible outside pilot #5 UM/UKCM. DTA with UKCM as data owner and UM as data processor will be prepared and signed prior to the trial.
5.2	5	PREMs/PROMs related to clinical staff	A GDPR compliant online survey infrastructure (https://1ka.arnes.si/index.php?lang_id=2)	SPSS or similar	JSON/CSV	Statistical data and cohorts available to be shared with the consortium
5.2	5	PREMs related patients	A GDPR compliant online survey infrastructure (https://1ka.arnes.si/index.php?lang_id=2)	SPSS or similar	JSON/CSV	Statistical data and cohorts available to be shared with the consortium
3.5	5	Datasets for facial expression and	At UM's internal infrastructure	N/A	N/A	The data is openly available, request must be issued by each

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
		emotion recognition				individual partner to the specific owner of the data set
3.5	5	Datasets for ASR and TTS in Slovenian	At UM's internal infrastructure	N/A	N/A	The data is not publicly open. Access can be granted on individual basis (bilateral agreements which may include charges)
3.5	2	Datasets for ASR and TTS in French	At UM's internal infrastructure	N/A	N/A	Access to public datasets must be managed by individual partners, access to UM's closed datasets can be granted on individual basis (bilateral agreements which may include charges)
3.5	5	EVA Corpus, data set of conversational expression	At UM's internal infrastructure And CLARIN.SI Repository	N/A	N/A	Access via CLARIN.SI repository

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
3.1 & 5.2	1	Cardiac ultrasound video recordings	At the edge, dedicated infrastructure for the pilot	TBD	DICOM, HL7-compatible resource	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Request must be issued by each individual partner to AHEPA.
3.1 & 5.2	1	Capsule endoscopy video recordings	At the edge, dedicated infrastructure for the pilot	TBD	MP4, HL7-compatible resource	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Request must be issued by each individual partner to AHEPA.
5.2	3, 6	IMU data captured by iPrognosis smartphone application	Cloud infrastructure for the pilots	Schemaless database	JSON, HL7-compatible resource	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
						pending publication or patent). Data access roles will be granted.
5.2	3, 6	Voice-related time and spectral features captured by iPrognosis smartphone application	Cloud infrastructure for the pilots	Schemaless database	JSON, HL7-compatible resource	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Data access roles will be granted.
5.2	3, 6	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	Cloud infrastructure for the pilots	Schemaless database	JSON, HL7-compatible resource	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Data access roles will be granted.

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
5.2	3, 6	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	Cloud infrastructure for the pilots	Schemaless database	JSON, HL7-compatible resource	Access to pseudo-anonymised data is allowed with no restrictions, unless otherwise decided (e.g., in case of pending publication or patent). Data access roles will be granted.
5.2	6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	Cloud infrastructure for the pilots	TBD	JSON	Under the conditions of the Data Processing Agreement between the partners involved in pilot #6.
3.1 & 5.2	1	Cardiotocography variables and results, biometric data, medical history data	At the edge, dedicated infrastructure for the pilot	TBD	TBD	It is not available/accessible outside pilot #1 AUTH.

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
3.1 & 5.2	1	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	At the edge, dedicated infrastructure for the pilot	TBD	DICOM partially (others TBD)	It is not available/accessible outside pilot #1 AHEPA/AUTH.
3.3	3	Electricity consumption and production, human presence/access control, devices activation and connected load	On VIMAR cloud infrastructure	InfluxDB	JSON	Data belongs to the owner of the location where the home automation devices are installed. The owner defines access to the data
5.3	PhE	All pilot data with Common KPIs (Economic & PROMs/PREMs)	The original data at each pilots' repository. The analysis of PhE will be stored in cloud space of PhE (One Drive)	R, Stata	Econometric data, *dta, *xlsx,	The primary data is property of each pilot, and the results of the economic/PROM/PREM analysis will be available to all partner. Coding of the

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
						economic part is property of PhE.
3.2	2	Data related to hours spent by specialists.	CHUL	N/A	TBD	None
3.2	2	Retrospective patient schedule data and precondition of the treatment	CHUL	Mosaiq OIS	TBD	Under the conditions of the Data Processing Agreement between the partners involved in pilot #2
1.2	2	Data linked to radiotherapy machines (tumour treatment indication, maintenance, building location)	CHUL	Mosaiq OIS	TBD	None

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
5.2	2	PROMs/PREMs and chatbot	CHUL	CHUL	TBD	PROM and PREM questionnaire no restriction, however TBD how chatbot data will be shared with partners
5.2	2	Prospective patient clinical data	CHUL	EHR	TBD	Under the conditions of the Data Processing Agreement between the partners involved in pilot #2
5.2	2	Patient personal data (address, preferences, …)	CHUL	EHR	TBD	Under the conditions of the Data Processing Agreement between the partners involved in pilot #2.

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
3.2 and 5.2	2	Retrospective EHR	CHUL	EHR	TBD	Under the conditions of the Data Processing Agreement between the partners involved in pilot #2.
3.5	6	Patient's physical, medical and mental status. Vital signs with sensors data.	At the edge, dedicated infrastructure for the pilot	Data Store (HBASE, MONGO, RDBMS)	Media Resource of HBASE, MONGO, RDBMS Observation	Private patient information is not available/accessible outside pilot #6. What we can share in extracted is features + pattern identification outcome since this is anonymized and will be also exploited in publications
3.2	2	Data from radiotherapy services, infrastructure. Patient's satisfaction	At the edge, dedicated infrastructure for the pilot	Data Store (HBASE, MONGO, RDBMS)	Media Resource of HBASE, MONGO, RDBMS Observation	Private patient information is not available/accessible outside pilot #2. What we can share in extracted is features + pattern identification outcome since this is anonymized and

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
						will be also exploited in publications
?	8	Image, gene, phenotype and pathology data for glioma patients	Separate in-hospital databases, with final central integration on site	PRIMUZ, XNAT	RDMS (various), JSON	The data is restricted, we will pursue ways to make the data (partially) available to the consortium via pseudo anonymisation; work is ongoing with UZ Brussel.
3.6	7	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	UZB, VUB	Electronic health records system branded PRIMUZ	xlsx files	Data will be not available/accessible outside pilot #7
3.6	7	Clinical data of patients scheduled for a coronary angiogram/coronar	UZB, VUB	Electronic health records system	xlsx files	Data will be not available/accessible outside pilot #7

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
		y intervention @ UZB		branded PRIMUZ		
3.6	7	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	UZB, VUB	Philips cathlabs over the Philips IntelliSpace CardioVascual (ISCV) Portal	DICOM	Data will be not available/accessible outside pilot #7
3.6	7	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and	UZB, VUB	Coroventis platform (RFR/FFR) and Volcano platform (iFR/FFR)	Data extracted in .dat or .xls files	Data will be not available/accessible outside pilot #7

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
		performed before and after a coronary intervention @ UZB				
3.6	7	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	UZB, VUB	Abbott OPTIS platform (OCT) (Integrated platform) Volcano platform (IVUS) (non-integrated platform)	TBD	Data will be not available/accessible outside pilot #7
3.6	7	Coronary CT data including FFRCT computation of patient referred for	UZB, VUB	Philips (CT) Heartflow (FFRCT)	TBD	Data will be not available/accessible outside pilot #7

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
		an invasive coronary angiogram and/or a coronary intervention				
3.6	7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	Philips AI smart-cathlab prototype working at UZB, VUB	TBD	TBD	Data will be not available/accessible outside pilot #7
	4				TBD	
3.5	5	Video recordings of the patient	Inside the robot hardware or attached devices	TBD	Bytestreams	Private patient information is not available/accessible outside pilot #5.

Task Nr.	Pilot Nr.	Type of data to be collected	Physical location where primary (original) data is stored	Software/tools used for storage of data?	Format and type of data standards used to store the data	How will the data be made accessible to other partners in consortium? Are there any restrictions?
3.5	5	Video recordings of third persons	Inside the robot hardware or attached devices	TBD	Bytestreams	Private patient information is not available/accessible outside pilot #5.
3.5	5	Patient's behavioural information	At the edge, dedicated infrastructure for the pilot	Data Store (HBASE, MONGO, RDBMS)	TBD	Private patient information is not available/accessible outside pilot #5.
3.5	5	Patient's facial information	At the edge, dedicated infrastructure for the pilot	Data Store (HBASE, MONGO, RDBMS)	TBD	Private patient information is not available/accessible outside pilot #5.

c. Please describe whether data are findable, including provisions for metadata

Task Nr.	Pilot Nr.	Type of data to be collected	Data identification and versioning	Naming conventions	Search keywords provided for re-use	Metadata
3.1 & 5.2	1	Cardiac ultrasound video recordings	TBD	TBD	TBD	TBD
3.1 & 5.2	1	Capsule endoscopy video recordings	TBD	TBD	TBD	TBD
5.2	3, 6	IMU data captured by iPrognosis smartphone application	TBD	TBD	TBD	TBD
5.2	3, 6	Voice-related time and spectral features captured by iPrognosis smartphone application	TBD	TBD	TBD	TBD
5.2	3, 6	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	TBD	TBD	TBD	TBD
5.2	3, 6	Joint coordinates of 3D skeleton data captured by the	TBD	TBD	TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Data identification and versioning	Naming conventions	Search keywords provided for re-use	Metadata
		iPrognosis iMAT application				
5.2	6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	TBD	TBD	TBD	TBD
3.1 & 5.2	1	Cardiotocography variables and results, biometric data, medical history data	TBD	TBD	TBD	TBD
3.1 & 5.2	1	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	TBD	TBD	TBD	TBD
5.3	PhE	Economic data from all pilots	TBD	TBD	AI economic evaluation	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Data identification and versioning	Naming conventions	Search keywords provided for re-use	Metadata
3.5	6	Patient's physical, medical and mental status. Vital signs with sensors data.	Age, Body Height, General status. Heartbeat. Oxygen, Blood pressure, Body temperature, Urine, Glucose, Bodyweight, Stool.	Heartbeat: X' Oxygen: % Blood pressure: mmHg Body temperature: ºC Urine: chromatic scale Glucose: mg-DL Bodyweight: Kg Stool: Number of times per day.	TBD	TBD
3.2	2	Data from radiotherapy services, infrastructure. Patient's satisfaction	TBD	TBD	TBD	TBD
Pilot #2	All data to defined					

Task Nr.	Pilot Nr.	Type of data to be collected	Data identification and versioning	Naming conventions	Search keywords provided for re-use	Metadata
TBD	8	Image, gene, phenotype and pathology data for glioma patients	Unique IDs via patient numbers, versioning for local systems exists, TBD for central one	TBD	TBD	TBD
3.6	7	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD
3.6	7	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD
3.6	7	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Data identification and versioning	Naming conventions	Search keywords provided for re-use	Metadata
3.6	7	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	TBD	TBD	TBD	TBD
3.6	7	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention	TBD	TBD	TBD	TBD
3.6	7	Coronary CT data including FFRCT computation of patient referred for	TBD	TBD	TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Data identification and versioning	Naming conventions	Search keywords provided for re-use	Metadata
		an invasive coronary angiogram and/or a coronary intervention				
3.6	7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	TBD	TBD	TBD	TBD
	4	TBD				
3.5	5	Video recordings of the patient	By robot ID and date	TBD	TBD	TBD
3.5	5	Video recordings of third persons	By robot ID and date	TBD	TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Data identification and versioning	Naming conventions	Search keywords provided for re-use	Metadata
3.5	5	Patient's behavioural information	By patient name or ID, robot ID and date	TBD	TBD	TBD
3.5	5	Patient's facial information	By patient name or ID	TBD	TBD	TBD

d. Please specify if data will be openly accessible and interoperable. Specify availability and usability by third parties, and licensing to third parties.

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
3.5	5	Datasets for facial expression and emotion recognition	YES	e.g., jaffe, FER-2013, MMI, Cohn-Kanade, RaFD, FERG, EMOTIC, Affect data, SemEval-2017 Task 4, DailyDialog, The MPLab GENKI Database, FFECTIVA-MIT Facial Expression Dataset (AM-FED), Grounded Emotions, Reldi, etc.	Licenses are granted on individual requests, not handled by UM but the owners of data sets	N/A	Licenses are granted on individual requests, not handled by UM but the owners of data sets

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
3.5	5	Datasets for ASR and TTS in Slovenian	NO				Access can be granted on individual basis (bilateral agreements which may include charges)
3.5	2	Datasets for ASR and TTS in French	NO				Access can be granted on individual basis (bilateral agreements which may include charges)
3.5	5	EVA Corpus, data set of	YES	Clarin.SI	N/A	N/A	CC-BY 4.0 License

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
		conversational expression					
5.2	1	Cardiac ultrasound video recordings	YES (unless otherwise decided due to IPRs)	Zenodo	TBD	HL7	TBD
5.2	1	Capsule endoscopy video recordings	YES (unless otherwise decided due to IPRs)	Zenodo	TBD	HL7	TBD
5.2	3, 6	IMU data captured by iPrognosis smartphone application	YES (unless otherwise decided due to IPRs)	Zenodo	TBD	HL7	TBD
5.2	3, 6	Voice-related time and spectral features captured by iPrognosis smartphone application	YES (unless otherwise decided due to IPRs)	Zenodo	TBD	HL7	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
5.2	3, 6	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	YES (unless otherwise decided due to IPRs)	Zenodo	TBD	HL7	TBD
5.2	3, 6	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	YES (unless otherwise decided due to IPRs)	Zenodo	TBD	HL7	TBD
5.2	6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	YES (but considering limitations described in the IPR/Consortium Agreement)	TBD	TBD	TBD	TBD
5.2	1	Cardiotocography variables and results, biometric	TBD	TBD	TBD	TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
		data, medical history data					
5.2	1	Coronary computed tomography angiography (CCTA) variables, biometric data, medical history data	TBD	TBD	TBD	TBD	TBD
3.3	3	Electricity consumption and production, human presence/access control, devices activation and connected load	Data belongs to the owner of the location where the home automation devices are installed. The owner defines access to the data.	None	The owner defines access to the data.	N/A	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
5.3	PhE	Economic Data and PROMs/PREMs of all pilots	Applicable for pilot specifics	Applicable for pilot specifics	Applicable for pilot specifics	Applicable for pilot specifics	Applicable for pilot specifics
3.5	6	Patient's physical, medical and mental status. Vital signs with sensors data.	NO	N/A	N/A	N/A	N/A
3.2	2	Data from radiotherapy services, infrastructure. Patient's satisfaction	NO	N/A	N/A	N/A	N/A
Pilot #2	All data to defined						
?	8	Image, gene, phenotype and	No	Not relevant	To be determined	To be determined	To be determined

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
		pathology data for glioma patients					
3.6	7	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD	TBD
3.6	7	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD	TBD
3.6	7	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	TBD	TBD	TBD	TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
3.6	7	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB	TBD	TBD	TBD	TBD	TBD
3.6	7	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention @ UZB	TBD	TBD	TBD	TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
3.6	7	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention @ UZB	TBD	TBD	TBD	TBD	TBD
3.6	7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	TBD	TBD	TBD	TBD	TBD
3.5	5	Video recordings of the patient	NO	N/A	N/A	N/A	N/A

Task Nr.	Pilot Nr.	Type of data to be collected	Are dataset made available openly?	Which open repository?	Licenses, access identification, restrictions	Metadata vocabularies, ontologies, standards and methodologies for data interoperability	Availability and usability of data by third parties. Licensing
3.5	5	Video recordings of third persons	NO	N/A	N/A	N/A	N/A
3.5	5	Patient's behavioural information	NO	N/A	N/A	N/A	N/A
3.5	5	Patient's facial information	NO	N/A	N/A	N/A	N/A

e. Please describe the process of cleansing, transforming and analyzing the data. Please be specific for each type of data.

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
5.2	1	Cardiac ultrasound video recordings	Add missing metadata, if applicable	AHEPA; person(s) TBD	Segmentation / Histogram normalization / Resolution adjustment / Machine learning model training & testing	Python/ TensorFlow / PyTorch / SciKit Learn	AUTH; person(s) TBD	TBD
5.2	1	Capsule endoscopy video recordings	Add missing metadata, if applicable	AHEPA; person(s) TBD	Segmentation / Histogram normalization / Resolution adjustment / Machine learning model training & testing	Python/ TensorFlow / PyTorch / SciKit Learn	AUTH; person(s) TBD	TBD
5.2	3, 6	IMU data captured by iPrognosis smartphone application	Remove duplicates / Remove corrupted/	AUTH; person(s) TBD	Pre-processing (sampling frequency adjustment, filtering, outlier detection) /	Python/ TensorFlow / SciKit Learn	AUTH; person(s) TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
			empty recordings		Input to and fine-tuning of pre-trained machine learning model that detects tremor			
5.2	3, 6	Voice-related time and spectral features captured by iPrognosis smartphone application	Remove duplicates / Remove corrupted/ empty recordings	AUTH; person(s) TBD	Input to and fine-tuning of pre-trained machine learning model that classifies user as PwP or not	Python/ TensorFlow / SciKit Learn	AUTH; person(s) TBD	TBD
5.2	3, 6	Key taps press and release timestamps captured by iPrognosis smartphone virtual keyboard	Remove duplicates / Remove corrupted/ empty recordings	AUTH; person(s) TBD	Input to and fine-tuning of pre-trained machine learning models that output bradykinesia and rigidity scores	Python/ TensorFlow / SciKit Learn	AUTH; person(s) TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
5.2	3, 6	Joint coordinates of 3D skeleton data captured by the iPrognosis iMAT application	Remove corrupted/ empty recordings	AUTH; person(s) TBD	Outlier detection/ Normalization / Feature extraction related to movement magnitude and speed / Similarity score to ideal movement	Python	AUTH; person(s) TBD	TBD
5.2	6	Results of rehabilitation sessions with Gradior and care plans (editor and patient data)	Remove duplicates, Remove Null values	INTRAS, AUTH, ITCL	Normalization	TBD	TBD	TBD
3.3	3	Electricity consumption and production, human presence/access control, devices	Outlier detection, inference on missing data	VIMAR	Normalization	TBD	TBD	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
		activation and connected load						
5.3	PhE	Economic data of all pilots PROMs/PREMs data	Data cleaning will be performed by each data owner (pilot). PhE will perform the analysis on clean datasets	TBD by each pilot.	Economic & PRO/PREM analysis (economic evaluation, cost consequence analysis, cost-utility analysis, patients' quality of life measurement, bootstrapping, regression, etc	Stata, Excel, maybe SAS if necessary	PhE	Following ISPOR guidelines for economic evaluation and Dolan's publication on EQ-5D analysis
3.5	6	Patient's physical, medical and mental status. Vital signs with sensors data.	Remove duplicates, Remove Null values	ITCL	Scale and normalize data.	Python libraries.	ITCL	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
3.2	2	Data from radiotherapy services, infrastructure. Patient's satisfaction	Remove duplicates, Remove Null values	TMA	Scale and normalize data.	Python libraries.	TMA	TBD
	Not applicable for CHU Liège							
	8	Image, gene, phenotype and pathology data for glioma patients	Connecting sample information per patient across databases, machine learning on data	To be hired	Integration, neural networks	Python, pytorch, sklearn, XNAT framework	VUB	To be determined

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
3.6	7	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Excel	Philips Image Guided Therapy Systems, Netherlands	TBD
3.6	7	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Excel	Philips Image Guided Therapy Systems, Netherlands	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
3.6	7	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Excel	Philips Image Guided Therapy Systems, Netherlands	TBD
3.6	7	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic index (FFR) either during a manual or a motorized wire pullback and	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low physiology curves quality)	Coroventis software and Virtual stenting algorithm (VSA) for interpretation of RFR/FFR pullback	Philips Image Guided Therapy Systems, Netherlands and JF Argacha, cardiology department,	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
		performed before and after a coronary intervention @ UZB				created by JF Argacha and Jean Decamp (I-depot number 123060 Date 16-04-2020)	UZB, VUB, Brussel	
3.6	7	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention@ UZB	Data cleaning will be performed by data owner (UZB)	UZB primary investigator will be responsible of data cleaning and preparation (TBD)	Manual removal of outlier patients (low image quality)	Optis and volcano software	Philips Image Guided Therapy Systems, Netherlands	TBD
3.6	7	Coronary CT data including FFRCT computation of	Data cleaning will be	UZB primary investigator	Manual removal of outlier patients (low image quality)	Philips and Heartflow software	Philips Image Guided	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
		patient referred for an invasive coronary angiogram and/or a coronary intervention@ UZB	performed by data owner (UZB)	will be responsible of data cleaning and preparation (TBD)			Therapy Systems, Netherlands	
3.6	7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary physiology, intravascular imaging and coronary CT data.	Not applicable (prospective inclusion)	Not applicable (prospective inclusion)	Not applicable (prospective inclusion)	Not applicable (prospective inclusion)	Philips Image Guided Therapy Systems, Netherlands and Cardiology department UZB, VUB	TBD
3.5	5	Video recordings of the patient	TBD	ITCL	TBD	Python libraries	ITCL	TBD

Task Nr.	Pilot Nr.	Type of data to be collected	Type of data cleaning needed (e.g. correct data types, remove duplicates, add missing infos…)	Person responsible for data cleaning (partner short name; person name; email)	Type of data transformation/analysis (e.g. normalization, discretization, …)	Software/tools used for cleaning, transform, and analyse	Where/by whom will the analysis be conducted?	Standards followed for code development / access and re-use
3.5	5	Video recordings of third persons	TBD	ITCL	TBD	Python libraries	ITCL	TBD
3.5	5	Patient's behavioural information	TBD	ITCL	TBD	Python libraries	ITCL	TBD
3.5	5	Patient's facial information	TBD	ITCL	TBD	Python libraries	ITCL	TBD

f. Please, describe how will you conduct quality control

Task Nr.	Pilot Nr.	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
5.2	1	TBD	TBD	AUTH; person(s) TBD	Peer-review
5.2	1	TBD	TBD	AUTH; person(s) TBD	Peer-review
3.3	3	TBD	TBD	TBD	TBD
3.5	5	TBD	TBD	ITCL	TBD
5.3	PhE	Based on Quality control Standard Operating Procedure (SOP) followed by PhE for all projects/deliverables	Protocol & CRF has been prepared by PhE and will be shared to all pilots in order to follow the same	PhE project team / Eugena Stamuli, Declan O' Byrne, Magda Chatzikou	TBD
3.2	2	TBD	Verification and validation plan	CHU de Liège, Patrick Duflot	System test, Unit test, Manual test
5.2	2	TBD	Monitoring Plan ensuring that the collected data are the right ones and completed.	CHU de Liège, Patrick Duflot	TBD in study protocol
5.3	2	As defined by PhE	As defined by PhE	As defined by PhE	As defined by PhE
3.5	6	Protocols shared by PhE	Protocols shared by PhE	ITCL	TBD
3.2	2	Protocols shared by PhE	Protocols shared by PhE	ITCL	TBD

Task Nr.	Pilot Nr.	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
?	8	Image, gene, phenotype and pathology data for glioma patients	To be determined in collaboration with UZ Brussels ICT	VUB; to be determined	Local testing, feedback from specialists, eventual peer review
3.6	7	Administrative data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
3.6	7	Clinical data of patients scheduled for a coronary angiogram/coronary intervention @ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
3.6	7	Coronary angiogram imaging data of patients who undergone a coronary angiogram/coronary intervention @ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
3.6	7	Coronary physiology data of patients evaluated by a resting index measure (iFR/RFR) or a hyperemic	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD

Task Nr.	Pilot Nr.	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
		index (FFR) either during a manual or a motorized wire pullback and performed before and after a coronary intervention @ UZB			
3.6	7	Intravascular imaging data of patient evaluated by either OCT or IVUS technique before and after a coronary intervention@ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
3.6	7	Coronary CT data including FFRCT computation of patient referred for an invasive coronary angiogram and/or a coronary intervention@ UZB	Retrospective study protocol (registry) is under preparation	Philips Image Guided Therapy Systems, Netherlands	TBD
3.6	7	Full prospective UZB data set of administrative, clinical, coronary angiogram, coronary	Prospective study protocol (comparative observational study) is under preparation	Primary investigator UZB, VUB (TBD)	Comparative analysis of different key performance indicators and

Task Nr.	Pilot Nr.	Type of analysis	Will you work according to specific protocol(s)? If yes, which one(s)?	Who will create the statistical analysis plan? (partner short name; person name; email)	How will data transformation & analysis be verified?
		physiology, intravascular imaging and coronary CT of data.			patient-reported outcome and experience measures (PREMS and PROMS) evolution before (control period) and after implantation of AI smartcathlab prototype (study period)
	4	TBD		91	

g. Governance and security: please complete <u>for each participating partner</u>

Classification degrees:
For the assessment of governance and security, the maturity of the process is classified in maturity levels using values that are inherited from the COBIT standardization guidelines22. Those values are:

I	**Initial**	Process not specified, based on spontaneous initiative, that is poorly controlled and reactive.
M	**Managed**	Process is planned, documented and monitored at the project level but not integrated in a broader scope at organization level.
D	**Defined**	Proactive process active at organization level.
Q	**Quantitively Managed**	The process is measured and controlled / verified.
O	**Optimizing**	Focus is on continuous process and improvement.

Partner short name	Classification degree	Code of conduct used for each task	What levels of data security do you have in place?	What levels of security will undergo primary data used in the project be? Please list all	Name and email of the data privacy officers in your organization
EXYS	Q	Data handling procedures at EXYS are defined according to GDPR and LPD (Swiss data protection law)	Technical measures: • Data storage in dedicated servers segmented on network environment	Whenever required by the project, data will be stored in firewall-protected computers with strong authentication.	Angelo Consoli, angelo.consoli@eclexys.com

22 ISACA®, COBIT® 2019 Framework: Governance and Management Objectives, USA 2018

Partner short name	Classification degree	Code of conduct used for each task	What levels of data security do you have in place?	What levels of security will undergo primary data used in the project be? Please list all	Name and email of the data privacy officers in your organization
			• Authorized access with data access audit logs • VPN regulated access to the data processing data center Data communication and transfer protected by Secure Socket Layer (SSL) and Transport Layer Security (TLS) protocols.		
AUTH		Data are handled according to GDPR and international ethical guidelines (inter alia, the Word Medical Association Declaration of Helsinki), mandated by the AUTH Research Ethics Committee. Ethical approval must be obtained before data collection	Technical measures include: • Data storage in firewall-protected computers • Authorized access with data access audit logs • Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols.	• Data will be stored in firewall-protected computers with authorized access. • Access will be limited to members of the AUTH research team. • In case of data transfer, this will take place via SSH or SSL protocols. • Pseudo-anonymised data will be stored separately from records	Ms. Kornilia Skarpeta, data.protection@auth.gr

Partner short name	Classification degree	Code of conduct used for each task	What levels of data security do you have in place?	What levels of security will undergo primary data used in the project be? Please list all	Name and email of the data privacy officers in your organization
		involving human beings.		including subjects' personally identifiable information (e.g., signed consent forms)	
VIMAR		Data are handled according to GDPR.	Technical measures include: • Data storage in firewall-protected computers. • Authorized access with data access audit logs. • Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols.	• Data will be stored in firewall-protected computers with authorized access. • Access will be limited to authorized members of the team. • In case of data transfer, this will take place via SSL protocols. • Pseudo-anonymised data will be stored separately from records including subjects' personally identifiable information.	Beni Luigi Gianesin, beni.gianesin@vimar.com
PhE		Data analysis will be handled according to GDPR following PhE Standard Operating	Cloud – One Drive	Data will be password protected and accessed only by the PhE HOSMARTAI dedicated team.	George Tsonis, Tsonislaw@yahoo.gr

Partner short name	Classification degree	Code of conduct used for each task	What levels of data security do you have in place?	What levels of security will undergo primary data used in the project be? Please list all	Name and email of the data privacy officers in your organization
		Procedure of Process & Handling of Personal Data.			
CHUL	Q	Information Security Plan	• Recruiting and training, • Security board, • Internal and external audit, • Vulnerability assessment, • Malware and virus protection, • Monitoring and alerting, • Incident management, • Encryption at rest and during transfer, • Confidentiality and access rights, • Asset management, • High availability, • Identity management, • 2-factors authentication,	Data protection impact assessment should define the technical and organizational measures to put in place.	dpo@chuliege.be

Partner short name	Classification degree	Code of conduct used for each task	What levels of data security do you have in place?	What levels of security will undergo primary data used in the project be? Please list all	Name and email of the data privacy officers in your organization
			• Single sign-on, • Backups, • Certifications: ISO27001, ISO9001, PRINCE2		
ITCL		Data are handled according to GDPR and international ethical guidelines. Ethical approval must be obtained before data collection involving human beings.	Technical measures include: • Data storage in firewall-protected computers • Authorized access with data access audit logs • Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols. • Encryption of data.	Data will be stored in firewall-protected computers with authorized access. Access will be limited to members of the ITCL research team. In case of data transfer, this will take place via SSH or SSL protocols. Pseudo-anonymised data will be stored separately from records including subjects' personally identifiable information (e.g., signed consent forms)	Angel Lopez, angel.lopez@itcl.es

Partner short name	Classification degree	Code of conduct used for each task	What levels of data security do you have in place?	What levels of security will undergo primary data used in the project be? Please list all	Name and email of the data privacy officers in your organization
				Critical data will be encrypted for better patient safety.	
VUB	I/M	Data are handled according to GDPR and international ethical guidelines, as well as internal UZ Brussel patient data security rules.	The data will remain in the tightly controlled and firewalled UZ Brussel ICT system, with only local or VPN access to limited authorised subsystems.	Limited access by authorized personnel only, no access from outside nor to outside from virtual machine harbouring the data	Data protection office, (dpo@vub.be)
UZB, VUB		Data will be handled according to EU 2016/679 GDPR.	Data storage in firewall-protected computers Authorized access with data access audit logs Data transfer via Secure Shell (SSH) or Secure Socket Layer (SSL) protocols.	Data will be stored in firewall-protected computers with authorized access. Access will be limited to authorized members of the team. In case of data transfer, this will take place via SSL protocols.	Luc Maes, lucm@uzbrussel.be
INTRAS	Q	Data handling procedures at INTRAS	Technical measures: • Data storage in dedicated servers	• Data will be stored in firewall-protected	Francisco Mallo Vázquez, fmv@intras.es

Partner short name	Classification degree	Code of conduct used for each task	What levels of data security do you have in place?	What levels of security will undergo primary data used in the project be? Please list all	Name and email of the data privacy officers in your organization
		are defined according to GDPR	segmented on network environment • Authorized access with data access audit logs • VPN regulated access to the data processing data centre Data communication and transfer protected by Secure Socket Layer (SSL) and Transport Layer Security (TLS) protocols.	computers with authorized access. • Access will be limited to members of the INTRAS research team. • In case of data transfer, this will take place via SSH or SSL protocols. Pseudo-anonymised data will be stored separately from records including subjects' personally identifiable information (e.g., signed consent forms)	