D1.8

Data Management Plan – V2

ABSTRACT

The goal of the InteropEHRate project is to provide secure cross-country interoperability of health records by means of a mobile and web-based Application (Device 2 Device and Web) without a central institution. In the course of the development of the various software applications and the corresponding administrative processes including the management of the project, data is produced and collected. This deliverable describes the FAIR (Findable, accessible, interoperable and reusable) management of these research data sets.

Delivery Date	July 7th, 2021 (resubmission on August 31st 2021)
Work Package	WP1
Task	T1.4
Dissemination Level	Public
Type of Deliverable	Report
Lead partner	UNIVIE

InteropEHRate project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826106.

CONTRIBUTORS

	Name	Partner
Contributors	Nikolaus Forgó	UNIVIE
	Tima Otu Anwana	UNIVIE
	Katerina Polychronopoulos	UNIVIE
	Marie-Catherine Wagner	UNIVIE
	Felix Zopf	UNIVIE
	Vincent Keunen, Lucie Keunen	A7
	Patrick Duflot	CHU
	Paul De Raeve	EFN
	Tino Marti	EHTEL
	Laura Pucci Francesco Torelli	ENG
	Salima Houta Marcel Klötgen	FRAU
	Stefano Dalmiani	FTGM
	Christina Kotsiopoulou	HYG
	Adrian Bradu	SIVECO
	Sofianna Menesidou	UBITECH
	Gábor Bella	UNITN
	Thanos Kiourtis	UPRC
	Chrysostomos Symvoulidis, Stella Dimopoulou	BYTE
Reviewers	Simona Bica	SIMAVI
	Laura Pucci Francesco Torelli	ENG

LOG TABLE

Version	Date	Change	Partner
0.1	1-09-20	Update of DMP Questionnaire	All Consortium Partners
0.2	01-2021	Update of FAIR data set tables based on the questionnaire responses.	UNIVIE
0.3	02- 2021 03-2021	Review and amendments	SCUBA, FTGM, CHU, HYG, A7, ENG, BYTE, SIMAVI, UBITECH, UPRC, UNITN

0.4	04-2021	Update of version 2 of DMP	UNIVIE
0.5	05-2021	First Internal review	UNIVIE
0.6	18-06-2021	Clinical Partners Review	FTGM, CHU
0.7	21-06-2021	Second Internal Review	ENG, UNITN
0.8	22-06-2021	Technical Review	ENG
0.7	30-06-2021	Submission of DMP Version 2	UNIVIE
VFInal	09-07-2021	Final check and submission	ENG
	31-08-2021	New version without Annex 3	UNIVIE

ACRONYMS

Acronym	Term and definition
A7	A7 Software
BYTE	BYTE Computer Anonymi Viomichanikiemporiki Etaireia
D2D	Device to Device protocol
CHU	Centre Hospitalier Universitaire De Liège
DMP	Data Management Plan
EC	European Commission
EFN	Fédération Européenne Des Associations Infirmières Aisbl
HER	Electronic Health Record
EHTEL	European Health Telematics Association
EMR	Electronic Medical Records
ENG	Engineering - Ingegneria Informatica SPA
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
FHIR	Fast Healthcare Interoperability Resources
FTGM	Fondazione Toscana Gabriele Monasterio Per La Ricerca Medica E Di Sanita Pubblica
FRAU	Fraunhofer (Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung E.V.)
GA	Grant Agreement
GDPR	General Data Protection Regulation
HCP	Health-Care Professional
HCP App	Health-Care Professional Application

HIS	Health Information Systems
HL7	Health Level 7
HYG	Diagnostikon Kai Therapeftikon Kentron Athinon Ygeia Anonymos Etaireia
IDE	Integrated Development Environment
MHD	Mobile access to Health Documents
N/A	Not available
NCPeH	National Contact for Points for eHealth
QA	Quality Assurance
R2D	Remote to Device protocol
RDDI	Research Definition Document Interface
RDD	Research Data Description
RDS	Research Data Sharing (protocol)
SCUBA	Spitalul Clinic De Urgenta Bagdasar-Arseni
SIMAVI	Software Imagination & Vision SRL
S-EHR	Smart Electronic Health Record
UBITECH	UBITECH Limited
UNITN	Universita Degli Studi Di Trento
UPRC	University Of Piraeus Research Center

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Scope of the Document

This Document is the second and month 30 final update of the InteropEHRate Data Management Plan (DMP), version one [D1.7] was submitted during the first year of the project. The DMP analyses the main elements of the InteropEHRate data management policy. It is intended to cover the complete life cycle of the research data created and processed. The DMP and will outline the following:

- The types of research data that will be generated or collected during the project;
- How the research data will be processed and preserved;
- Which parts of the datasets will be shared for verification or re-use;
- The standards that will be used;
- The handling of research data after the end of the project.

The DMP aims to monitor the privacy and confidentiality of data, as well as ensure that the legal and ethical standards for data generation, use, storage, exchange and share are applied throughout the project. The DMP is one of the means by which the InteropEHRate Consortium ensures that data is processed in accordance with the provisions in the Grant Agreement, Consortium Agreement and applicable Regulations and standards. Furthermore, the DMP seeks to ensure that InteropEHRate activities are compliant with the Guidelines on "Implementation of Open Access to Scientific Publications and Research Data'' [ERC]. The DMP will address the measures employed by the Consortium partners to fulfil legal, ethical and privacy requirements concerning personal data usage and to assure compliance with relevant national or EU regulations, primarily the General Data Protection Regulation [GDPR].

In this document, organisations participating in the InteropEHRate project will be referred to as "partners" and all partners will collectively be referred to as the "Consortium". The InteropEHRate Consortium commits to respect the policies outlined in this DMP and ensure that all data is created, managed and processed in accordance with applicable legislation. The partners that generate or collect data are in charge of its integrity, compatibility, backups, validation and registration during the lifetime of the project. Backing up data for sharing through open access repositories is within the responsibility of the partner processing the data. All partners chairing a lead role for a specific project task outlined in the Project Grant Agreement have to assume responsibility for the quality control of the data generated or processed during the work on that specific task.

The information required to produce the initial DMP was gathered by the InteropEHRate Consortium through a questionnaire (see ANNEX 1 which follows the "Guidelines on FAIR Data Management in Horizon 2020"). The questions concerning various aspects of data management were taken from these guidelines and were arranged according to the template provided by Horizon 2020. The overall structure of this DMP is also based on this template. The tables depicted in Section 4 and 5 of this document represent mappings of this FAIR Data Management model. This new version of the DMP reflects the current approach to data management adopted in the Project and was developed based on the updated answers and reviews from each partner to the new Questionnaire (ANNEX 1) circulated at the end of the second year of the project.

1.2 Intended Audience

This DMP is a public document. The deliverable will be publically available on the InteropEHRate website (https://www.interopehrate.eu/). This document will primarily be interesting for health institutions (hospitals, clinic centres), health care personnel, health research communities, health magazines, Small and Medium Entrepreneurs (SMEs) or web entrepreneurs producing/implementing apps in the health domain as well as for potential secondary users of the data. Furthermore, this document is relevant for all who participate in the InteropEHRate pilot and validation phase. This includes Healthcare Organisations, Healthcare Practitioners, Researchers and Patients.

1.3 Structure of the Document

After an introduction covering different aspects of the InteropEHRate DMP and an executive summary of the project, a brief evaluation of FAIR Data Management will follow, including evaluations of the project data with respect to the various aspects of a life cycle of research data and data quality. Then a general survey of the data, generated and collected in the course of the project, is presented. The following chapters cover a more detailed analysis of the life cycles of the various data sets.

1.4 Updates with respect to Previous Version

This document updates and supersedes the previous version [D1.7] "D.1.7 - Data Management Plan V1". The updates to this document are based on new information provided by Consortium Partners concerning data management in the project. The following table provides an overview of the sections that have been updated and the contributions provided by each partner.

Section	Update	Partner
Section 2: Methodology	The methodology adopted to gather information for version 2 of the DMP has been included.	UNIVIE
Section 3: Synopsis "FAIR" Interoperate data sets	The FAIR principles table, which presents a concise survey of the key features of the life cycle of data, has been updated. These updates are based on contributions from Consortium Partners.	UNIVIE, ENG, A7, UBIT, SIMAVI

Section 4: Process-oriented analysis of "FAIR" data sets	The following sections have been updated to reflect the current state of data management in the project: Section 4.1 focusing on the management of administrative data is updated based on the contributions of ENG. Section 4.2.1 focusing on the management of technical data concerning cross-border electronic health records (EHRs) has been updated based on the contributions of UNITN. Section 4.2.2 focusing on the management of technical data concerning the S-EHR mobile app has been updated based on the contributions of A7. Section 4.2.3 focusing on the management of technical data concerning the HCP app has been updated based on the contributions of SIMAVI. Section 4.2.4 focusing on the management of technical data concerning interoperability has been updated based on the contributions of UPRC. Section 4.2.5 focusing on the management of technical data concerning the interoperability profile and standardization has been updated based on the contributions of FRAU. Section 4.2.6 focusing on the management of technical data concerning health record security and privacy has been updated based on the contributions of UBITECH.	UNIVIE, ENG, UNITN, A7, SIMAVI, UPRC, FRAU, UBITECH
Section 4.3: Pilot Project	This section contains new information regarding processing of data and data management in the context of the InteropEHRate pilots. This section is developed in collaboration with clinical and technical partners involved in the pilots: Section 4.3.1 on the Initial S-EHR Feed (Pilot 1, 2 and 3). Section 4.3.2 Pilot 1: Citizen Centred Healthcare Pilot. Section 4.3.3 Pilot 2: EHRs Integration for Emergency Pilot. Section 4.3.4 Pilot 3: Citizen Centred Medical Research Pilot. Section 4.3.5 Anonymization Approach.	ENG, UNITN, A7, BYTE, SIMAVI, UPRC, FRAU, UBITECH, FTGM, CHU, HYG, SCUBA. Section 5.3.5 - input made by BYTE
Section 5: Data Protection and Ethical Aspects	Section 5.1 The Ethical Package. This section is included to explain the package of legal and ethical documents, which will be checked and approved by ethical committees prior to the commencement of the pilots. The following new sections have been included: Section 5.1.1 Declaration of compliance. Section 5.1.2 Data Protection Impact Assessment. Section 5.1.3 Data Sharing Agreement. Section 5.1.4 Consent.	UNIVIE

Table 1- Updates with Respect to Previous Versions.

2 METHODOLOGY

2.1 FAIR Management of Research Data

The Miriam Webster Dictionary defines data as "information in digital form that can be transmitted or processed".[1] The International Organisation for Standardisation (ISO) defines data as "recorded information" and describes data management as "process of keeping track of all data and/or information related to the creation, production, distribution, storage, […] use of e-media and associated processes" [ISO 20294].

The InteropEHRate Data Management follows the "Guidelines on FAIR Data Management in Horizon 2020" [FAIR], released by the European Commission Directorate – General for Research & Innovation. According to these guidelines, the management and organization of data should be based on four basic principles. These principles determine how research outputs should be processed so that they can be more easily accessed, understood, exchanged and reused. This means that data must be findable, accessible, interoperable and reusable. These principles provide guidance for scientific data management and are relevant to all stakeholders in research projects. The guidelines directly address data producers and data. Research libraries can use the FAIR Data Principles as a framework for fostering and extending research data services. The different aspects of the use of data during a research process are depicted in Figure 1 below:

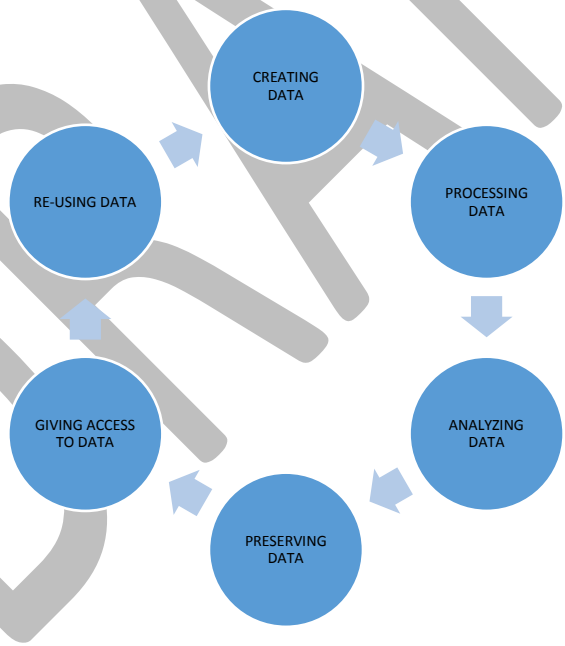

Figure 1 - Research Data Lifecycle [DATA ARCHIVE]

Creating data: designing research DMPs, planning consent, locating existing data, data collection and management, capturing and creating metadata.

[1] https://www.merriam-webster.com/dictionary/data

Processing data: entering, transcribing, checking, validating and cleaning data, anonymizing data, describing data, managing and storing data.

Analysing data: interpreting & deriving data, producing outputs, authoring publications, preparing for sharing.

Preserving data: data storage, back-up & archiving, migrating to best format & medium, creating metadata and documentation.

Access to data: distributing data, sharing data, controlling access, establishing copyright, promoting data.

Re-using data: follow-up research, new research, undertake research reviews, scrutinizing findings, teaching & learning

2.2 The Quality of Research Data

The quality of data is essential for an efficient and successful performance of the InteropEHRate platform and is critical to potential re-use of the EHRs for future research purposes. According to ISO/IEC 25024 (2015), quality of data is defined as the "degree to which the characteristics of data satisfy stated and implied needs when used under specified conditions" [ISO 25024].

There are six components that ensure data quality: completeness, consistency, accuracy, validity, timeliness, and uniqueness. The proper execution of each of these components will result in high quality data. The quality of data can be determined by these characteristics, which can be checked during a data quality assessment:

DATA QUALITY ASSESSMENT [DAMA]					
COMPLETENESS	CONSISTENCY	ACCURACY	VALIDITY	TIMELINESS	UNIQUENESS
The proportion of stored data against the potential of "100% complete".	The absence of difference, when comparing two or more representations of a thing against a definition.	The extent that data are correct, reliable, and certified free of error.	Data is valid if it conforms to the syntax (format, type, range) of its definition.	The degree to which data represent reality from the required point in time.	Nothing will be recorded more than once based upon how that thing is identified. It is the inverse of an assessment of the level of duplication.

Table 2- Data Quality Assessment

2.3 Collection of Information for the InteropEHRate Data Management Plan

The information for the DMP was gathered by the InteropEHRate Consortium through a questionnaire (see ANNEX 1), which is based and developed in accordance with the "Guidelines on FAIR Data Management" [FAIR]. Considering the different aspects of these recommendations, the project management and the partners agreed on a procedure used to map the specific details of the InteropEHRate project to the general structure provided by Horizon 2020.

The following table gives a survey which partner provided information on which data processes:

Collection of Information		
Data Process	Subsection	Partner (provided information for this DMP)
Information on Administrative Data (Section 4.1)	-	ALL
Information on Technical Processes (Section 4.2)	Cross-Border Electronic Health Records (EHRs) (Section 4.2.1)	UNITN
	Smart Electronic Health Records (S-EHR) Mobile App (Section 4.2.2)	A7
	Health Care Professional (HCP) App (Section 4.2.3)	SIVECO/SIMAVI
	Interoperability (Section 4.2.4)	UPRC
	Interoperability Profile and Standardization (Section 4.2.5)	FRAU
	Health Record Security and Privacy (Section 4.2.6)	UBITECH
Project Pilots (Section 4.3)	Initial S-EHR Feed (Pilot 1,2 and 3) (Section 4.3.1)	CHU, FTGM, ENG
	Pilot 1: Citizen Centred Healthcare Pilot (Section 4.3.2)	CHU FTGM, HYG, ENG, A7, UBITECH,
	Pilot 2: EHRs Integration for Emergency Pilot (Section 4.3.3)	SCUBA, FTGM, CHU
	Pilot 3: Citizen Centred Medical Research Pilot (Section 4.3.4)	FTGM, CHU

	Anonymization Approach (Section 4.3.5)	BYTE
Data Protection and Ethical Aspects (Section 5)	Ethical Package (Section 5.1)	ALL

Table 3 – Collection of Information

The information provided was transferred to the tables presented in the following sections.

The DMP is a living document. At the end of the 1st year of the project lifetime, all partners of the InteropEHRate Consortium reviewed and updated the information provided by them for the first version. No significant changes were reported. After another year (end of Year 2 of the project), a new Questionnaire (ANNEX 1) was compiled by the Partners to evaluate whether any significant changes occurred. At the beginning of Year 3 Partners reviewed their respective sections again. This second iteration of the DMP reflects the current approach to data management in the Project, specifically in the context of the upcoming pilots, i.e. the experimental and validation phase of the project where the InteropEHRate communication protocols and tools will be assessed at clinical pilot sites, by final users (patients, researchers and healthcare practitioners). The data management policy to be implemented at the pilot sites is detailed in Section 4 of this document. Further information on the experimentation and validation plan is outlined in deliverable 7.1 [D7.1].The effective and continued implementation of the DMP is part of a holistic and long-term data strategy.

3 SYNOPSIS: "FAIR" INTEROPERATE DATA SETS

The InteropEHRate Project will produce a variety of data. Most of this data falls into one of the following three major groups of data: administrative data (personal[2] and non-personal data), technical data (software code and design), and patients' data in pilots.

The following tables present a concise survey of the key features of the life cycle of this data. The information was derived from questionnaires filled out by relevant partners of the InteropEHRate Consortium (A7; CHU; EFN; EHTEL; ENG; FRAU; FTGM; HYG; SIVECO/SIMAVI; UBITECH; UNITN; UPRC). The subsequent sections will provide a more detailed analysis of the generating processes for various data and the resulting characteristics of each group. The following tables provide additional details about each of the three content related groups. In accordance with the FAIR principles - already addressed - specific information on data created/collected and processed in the respective processes is provided here:

	DATA PRODUCTION AND STORAGE
Data Generated/Collected	Administrative Data: Reports/Deliverables, Templates, Mailing lists, Partner Contact Details and Meeting related materials (participants' lists, attendees' signatures, agendas and presentations). Technical Data: Software Code, Software Design Patient Data in Pilots: Identification Data, Contact Data, Consent, Health Data (detailed in Section 4.3 Project in Pilots)
Data Format	Administrative Data: PDF, ZIP, Google Docs, Google Sheets, Google Slides, Excel, PowerPoint Presentations, Videos and images. Technical Data: JavaScript files Patient Data in Pilots: Paper documents, HL7-V4-RIM and HL7-FHIR
Reproducibility	Administrative Data: Google Drive and local repositories Technical Data: GitLab and local repositories Patient Data in Pilots: No
Size of the Data	Administrative Data: No larger than a few Gigabytes Technical Data: Approximately 1 Mb Patient data in Pilots: Pilot 1: 10-100 Mb, Pilot 2: 100-500 Mb depending on the data available on the S-EHR Cloud, Pilot 3: 10-500 Mb.
Software tools for creating/processing /visualizing data	Administrative Data: Google Drive Technical Data: IDE Patient Data in Pilots: S-EHR App, HCP App, S-EHR Cloud, Emergency identity token, EHR systems of Research Centres (FTGM & CHU), InteropEHRate Research Services, InteropEHRate Data Integration Platform (detailed in Section 4.3)

[2] All partners of the project agreed to have their personal data processed for administrative purposes for the lifetime of the project.

Use of pre-existing data	Administrative Data: No Technical Data: Open Source libraries, GitLab of standardization projects, Wiki of Standardization projects, pre-existing data from coding lists, nomenclatures Patient Data in Pilots: Data from Hospitals' EHRs
Storage and Backup Strategy	Administrative Data: Google Drive, local repositories and EC Portal Technical Data: InteropEHRate GitLab, ISO 27001 Patient Data in Pilots: In accordance with each partners Hospitals policies (detailed in Section 4.3)
	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Standards for Documentation and Metadata	Administrative Data: No particular standards. Technical Data: This depends on the software in question; please see the tables in section 4.2 for more details. Patient Data in Pilots: Not Applicable.
Best Practices/Guidelines for Data Management	Administrative Data: Defined in the InteropEHRate project deliverable D1.5. Technical Data: Defined in D2.10. Patient Data in Pilots: EMA Guidelines for Good Clinical Practice (23/07/2015), Helsinki Declaration, ISO 27001 and ISO 9001.
Tools for Formatting Data	Administrative Data: No automatic tools currently used Technical Data: Android Studio, IDE Tool, Validation tools for technical specifications Patient Data in Pilots: will be used data formatted by S-EHR app and HCP App.
Directory and File Naming Convention	Administrative Data: Defined in the InteropEHRate Project deliverable D1.5. Technical Data: Java naming convention. Patient Data in Pilots: not Applicable for local files.

	DATA ACCESS
Risks	Administrative Data: unauthorized access. Technical Data: Stealing of GitLab credentials and access to software code by an unauthorized person. Patient Data in Pilots: unauthorized access.
Risk Management	Administrative Data: User and Password controls and by other specific security policy granted by Google Drive terms of usage. Technical Data: ISO 27001, Private GitLab repository (User and Password protection) Patient Data in Pilots: Encryption, anonymization, logical access control and

	traceability (detailed in Section 4.3).
Correct execution of the Access process	Administrative Data: ENG is responsible for the concrete execution of the access protocols. Technical Data: ENG Patient Data in Pilots: Each Pilot Leader and principal Investigator (FTGM, CHU, HYG and SCUBA)
Procedures to Follow a Data Breach	Administrative Data: Governed by Google policy for data security and management of data scenarios (detailed in section 4.1). Technical Data: The procedure for each set of technical data is detailed in section 4.2 of this document. Patient Data in Pilots: In accordance with the data breach protocol at each hospital pilot site and research centre (FTGM, CHU, HYG, SCUBA) (detailed in Section 4.3)

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Re-Use of Data	Administrative Data: public deliverables and publications will be publicly available via the InteropEHRate website and other public sources. Technical Data: Public deliverables containing technical information will be published on the InteropEHRate website. Some software code will be open source. Patient Data in Pilots: Only anonymized results of clinical research will be available for reuse.
Organization/ Labelling of Data for Easy Identification	Administrative Data: N/A Technical Data: Not defined yet Patient Data in Pilots: Anonymized data
Data Sharing Requirements	Administrative Data: Public deliverables are openly accessible on the project website. Other project administrative data is shared internally between Partners and on the EC portal; this data is not publicly shared. Technical Data: IDE, Standardization tools Patient Data in Pilots: No health data and other personal data will be re-used or shared between hospitals or with third parties. Anonymous answers to validation questionnaires will be re-used and published in scientific publications and disseminations as determined by the project.
Audience for Reuse	Administrative Public deliverables are available to any interested party (software vendors, healthcare organisations, public and private institutions). Technical Data: Anyone (software vendors, Healthcare organizations, research organizations, public and private institutions) Patient Data in Pilots: Anonymized data of the validation result may be re-used

	by researchers, research institutions and healthcare practitioners.
Restrictions on Re-Use of Data	Administrative Data: confidential deliverables as stated in Part A of the Grant Agreement are not publicly accessible. Consortium deliverables are published under a creative common license. Technical Data: Not for public deliverables Patient Data in Pilots: Beyond the scope of the pilots, health data and other personal data will not be re-used for purposes not compatible with the initial purposes of the treatment.
Publication	Administrative Data: InteropEHRate project website (www.InteropEHRate.eu) Technical Data: GitLab, scientific conferences, standardization workshops, InteropEHRate Website Patient Data in Pilots: Anonymized data of the validation results may be published in public deliverables on the InteropEHRate website and in scientific publications.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	Administrative Data: Google Drive, EC Portal and local repositories Technical Data: GitLab and back-up Patient Data in Pilots: Patients consent forms will be archived (in paper and digital form). Patients' consent forms and health data needed for further medical care is archived in the hospital's systems, in accordance with national legislation (further details contained in Section 4.3).
Data Retention	Administrative Data: anticipated to be for at least 5 years Technical Data: at least 5 years Patient Data in Pilots: Patients consent forms and health data needed for further medical care is retained in the hospital's systems, in accordance with national legislation (further details contained in Section 4.3).
File Formats	Administrative Data: .docx, .xls Technical Data: GitLab on Kotlin (.kt), Java files Patient Data in Pilots: Data collected through Hospital EMR and paper documents/form, in proprietary and HL7-V3-RIM format. Data collected through S-EHR will be in proprietary and HL7-FHIR format [FHIR].
Data Archives	Administrative Data: EC Portal Technical Data: Local repositories Patient Data in Pilots: According to national regulation (detailed in Section 4.3).
Long-term Maintenance of Data	Administrative Data: The Project Coordinator will archive the deliverables and project related material (minutes of meetings, agendas, presentations) for five

	years after the end of the project. Technical Data: Not yet defined, will be presented in the next version of the DMP Patient Data in Pilots: The hospital (HYG, FTGM, CHU, SCUBA) IT department and the local principal investigator will maintain data in the long-term.

Table 4 – Synopsis of InteropEHRate Data Sets.

4 PROCESS-ORIENTED ANALYSIS OF "FAIR" DATA SETS

This section addresses the life cycle of administrative data and standards of project documentation. It also provides descriptions of the major technical processes put in place for the InteropEHRate project and the data deriving from them. The tables show how the partners in charge of respective tasks deal with the data sets that are generated in the course of their work.

4.1 Administrative Data

A variety of administrative data is generated and collected during the course of the Project. Some examples are:

- Planning data, concepts, the work for planning consortium meetings, when and how to communicate with other partners to ensure deadlines for the project, etc.;
- Data from administrative and financial management;
- Deliverables and reports;
- E-mails and minutes, documentation of communication among members of the project.

Each deliverable, presentation and meeting minutes is edited by using the InteropEHRate document templates available in the "Templates" folder of the project official repository. Templates are available both in the shared Google Document format and in .docx/.pptx or .odt/.odp formats.

As the coordinator of the project, ENG is the partner tasked with managing administrative data. All partners produce administrative data through the generation of deliverables, reports, publications and other documents related to project management.

The following table provides a summary of the characteristics and standards to be followed with respect to administrative data generated and processed during the Project. Further information is documented in deliverable D1.5.

	DATA PRODUCTION AND STORAGE
Data Generated/Collected	Reports/Deliverables defined in the GA Templates Partner contact information Physical meetings/Web conference conducted via Google meet or GoToMeeting and related material (participants' lists, attendees' signatures, agendas, presentations)
Data Format	Google Documents PDF, Doc, Excel, PowerPoint Presentations, Videos and Images.
Reproducibility	Google Drive maintains a history of the documents edited through google tools. All the deliverables will be uploaded to the EC Participant Portal- Funding and Tender website (https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/home). The process is not replicable but several copies of the deliverables are produced by the partners. With respect to

	contractual documents, copies are maintained on the Participant Project site.
Software tools for creating/processing /visualizing data	Deliverables, meetings/web conference related material, effort data and contact details of partners can be visualized online through Google Drive or offline through usual document readers.
Storage and Backup Strategy	Google Drive, Project website and EC Portal. Living versions of deliverables are stored in the collaborative workspace of the project repository on the Drive of the Google Workspace Business Standard (Google workspace, formerly GSuite, is an integrated service offered by Google to manage documents purchased by the Coordinator for the InteropEHRate project), in the related WP/Task subfolders. Final and official submitted versions of deliverables are stored by the PC in the "Submitted Deliverables" folder of the project repository. All partners are required to upload their deliverables in these folders before sending them for internal review. Public deliverables are uploaded also to the InteropEHRate project website, after being officially submitted to the EC (www.interoperate.eu).

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Standards for Documentation and Metadata	pdf and doc format for reports
Best Practices/Guidelines for Data Management	The Project repository provided through the Drive of Google Workspace Business Standard is a collaborative space that is the main exchange means of documents among partners.
Tools for Formatting Data	No automatic tool. Validation by the QA Manager and the Project Coordinator.
Directory and File Naming Convention	• Naming of the document in each deliverable: <ProjectName>_<Document_number>-<Document_Name>_<v.#> • Naming emails: [IEHR][WP_number][Task_number]
Project and Data Identifiers	Deliverables names/identifiers are agreed with the EC and stated in GA
Automatic Creation of Metadata	No

	DATA ACCESS
Risk Management	Risk: • Deletion and/or modification • In the case of confidential data (confidential deliverable, contact details of partners, etc.) the risk is also of non-authorized access. On the Project repository, data is protected by user and password controls and by other specific security protocols granted by Google Drive's terms of usage. This protects any documents on the Repository from possible external damaging voluntary attacks that could delete or modify them. The risk of unforeseen permanent deletion of documents or folders by users having authorized access to the Repository is very low. Any document or folder stored on the Google Workspace is not deleted immediately from the Google Drive, actually it is moved to the Google bin, and is deleted only if the user empties the trash or automatically after 30 days. The Google Workspace administrator, whose account is managed by the Project coordination team, can recover the deleted data within 25 days after the trash was emptied [GOOGLE]. In addition, each final version of the deliverables that has been submitted to the EC is permanently stored to the EC participant Portal and can be downloaded at any time by any partners having an authorized account to the EC Portal. The risk of unforeseen modification of documents by users having authorized access to the Repository is very low because each google document maintains a history of all modifications to that document since its creation and previous versions of a document can be easily restored.
Data Access	Only specific persons indicated by the Partners Project Managers (including their deputies) to whom the Project Coordinator grants access can access data (on the Project repository).
Correct execution of the access process	Project Coordinator will be responsible for ensuring secure access protocols for administrative data.
Procedure to address the possibility of a data breach	Security of documents stored through Google Workspace Business purchased for InteropEHRate is granted by the Google policy for data security and management of data incidents. Google specifies that specific mitigation actions are put in place in a precise process to address any potential incidents affecting the confidentiality, integrity, or availability of customers' data [GOOGLE].

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Data Sharing Requirements	Public deliverables are openly accessible on the project website. Other Project administrative data is not intended to be publicly shared or otherwise made

	available to third parties.
Audience for Reuse	Public deliverables are intended for anyone (Software vendors, healthcare organizations, research organizations, public and private institutions) who wants to study, implement or experiment with the new protocols and architecture developed by the project for the exchange of health data by means of the citizen mediations.
Restrictions on Re-Use of Data	Consortium deliverables are published under a Creative Common License. InteropEHRate Deliverables are for public, restricted or confidential circulation, as stated in the Part A of the Grant Agreement.
Publication	The InteropEHRate project website makes available public information about the Project and the Consortium, disseminating objectives and outcome of the research to the general public and acting as a pathway for interested users to go deeper into details about project outcomes through the "contact us" section. The website is available at the following URL[3] www.InteropEHRate.eu. Public deliverables are published on the project website as soon as the EU Commission approves them. Public deliverables may also be published as draft documents on the project website after being sent to the Commission. Released public software will be accompanied by reports following the same publishing process. The report will contain the URL and instructions for obtaining and using the specific software.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	Deliverables and effort/financial data are stored in the EC Portal. The Project Coordinator will archive all deliverables and project related documentation on its internal storage for future audits until five years after the end of the Project.
Data Retention	At least 5 years after the end of the project. All project management data (deliverables and Project Progress Reports) uploaded on the EC Portal are permanently retained.
File Formats	The Project Coordinator archives its copies of deliverables in the original format. All final documentation is in pdf format.
Data Archives	The EC Portal is the archive location for project management data (deliverables and effort data reported in the Project Progress Reports). The EC Portal is considered an institutional archive that preserves the deliverables and other information submitted permanently.

[3] Both the www.Interopehrate.eu and www.Interoperate.eu domains have been reserved for the project website.

	According to the InteropEHRate Grant Agreement, each Partner is obligated to independently archive any documentation related to financial and technical activities conducted during the project for at least 5 years after the end of the project. The purpose is to demonstrate that costs claimed in the course of the project are eligible.
Long-term Maintenance of Data	The Project Coordinator will archive the deliverables and project related material (minutes of meetings, agendas, presentations) for five years after the end of the project.

Table 5- FAIR Lifecycle of Data Sets: Administrative Data

4.2 Technical Data

This section addresses the main software development data generated during the project, including the software code. After a brief description of the goals and the intended products, the life cycles of the data are discussed. The tables are based on the information that is available and decisions on the processing of data that have been made at this stage in the project. Some changes and updates are expected as the project continues to develop.

4.2.1 Cross-Border Electronic Health Records (EHRs)

4.2.1.1 *Objectives*

It is an essential goal of this project to develop an open, decentralized and scalable architecture for cross-border EHRs, which is interoperable with existing systems and infrastructures and supports structured (HL7 FHIR) and unstructured binary data [FHIR]. A fundamental characteristic of the HL7 FHIR standard is its extensibility: the data model can be extended with new fields with a view to reusability in different applications [FHIR]. The InteropEHRate project will develop a set of tools to dynamically map and translate data from legacy clinical database models to the InteropEHRate FHIR profile. These profiles will be based on international standards for clinical terminologies (e.g., ICD-10 for classification of diseases) for structured content, and on other relevant standards (e.g., DICOM for medical imaging) for unstructured data [FHIR]. This architecture supports different levels of interoperability: from a low level for the secure exchange of non-standard data, to higher levels where data is translated to a common HL7 FHIR profile and into the language expected by the consumer [FHIR].

4.2.1.2 *Software Development*

During the InteropEHRate Project, a health data integration platform is developed to support the adoption of InteropEHRate FHIR profiles. This back-end platform includes healthcare knowledge and related tools, which are organized in linguistic, terminological and ontological layers. They formally represent lexical units, schemas, and encoding standards used by member countries. A tool for the definition of mapping rules will be developed, so that health record data, which can be used by local services (such as a hospital), and Europe-wide for cross- country data exchange.

A service for automatic conversion of legacy systems data (EMR/EHR/HIS) to a FHIR profile will be designed. The Project will also produce software for the automatic extraction of structured knowledge from natural language contained in health records. This includes quantities and units of measure (e.g., from prescriptions) as well as information on diseases and medical interventions. Another example is that the software will be capable of automatically translating the knowledge extracted from health records into the natural language that is comprehensible to the users of the data, in a country other than the one of the data source.

In the course of the above-mentioned different software development processes data is produced. The life cycle of this data is documented on the following pages. “FAIR” principles are applied.

4.2.1.3 *FAIR Lifecycle of Data Set for Software Coding and Development of Cross-Border EHRs*

The following tables further address the data produced in connection with the creation of Cross-Border EHRs.

	DATA PRODUCTION AND STORAGE
Data generated/collected by UNITN	1. Software code 2. Data concerning general health knowledge (coding standards for diseases, drugs laboratory codes, FHIR schemas) 3. Synthetic (fake) patient EHR data generated and provided by clinical partner FTGM.
Data Format	1. Software code for the implementation of data integration/interoperability systems is written in Java. 2. General medical knowledge, such as medical encoding standards (e.g., ICD-10) and mappings between standards, will be converted to an internal multilingual and logical representation. 3. Initially in HL7 CDA format, they will be converted by the InteropEHRate logic to FHIR format to be used for system testing purposes.
Reproducibility	All processing concerning medical knowledge will be fully automated. Therefore, it is reproducible if the same input is given. If its generation is manual, the user will store the results. Software code data will additionally be stored within the GitLab versioning system provided by the project partner responsible for system integration. Documentation will be stored on Google Drive. Thus, an efficient backup strategy is provided.
Size of the Data	Data concerning Health knowledge - 100MB range, growing by 20%-30% each year.

	Synthetic EHR data - 100KB range growing by 100% each year. Software code (software code) - 1MB range, growing by 100% each year.
Software tools for creating/processing/ visualizing data	IEHR Data Mapping Tool, IEHR Knowledge Management Tools, SCROLL NLP tool, Data Integration Platform.
Use of pre-existing data	Pre-existing general medical knowledge (non-personal) will be standardized according to local and international standards and published online in this form. Pre-existing software code will also be used, either brought in as background or open-source third-party tools.
Storage and Backup Strategy	Versioning systems (both local and project-wide) will be used for non-personal data and knowledge, Google Drive for documentation data. Software code stored in GitLab repositories.

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Standards for Documentation and Metadata	The responsible partner (UNITN) uses best practices of the Software development community. Inline comments in software code, GitLab standard READMEs, project documentation as Google Docs.
Best Practices/Guidelines for Data Management	Standard Java development conventions and frameworks will be used (e.g., Maven).
Tools for Formatting Data	Standard Oracle Java formatting will be applied.
Directory and File Naming Convention	Files are stored in the official InteropEHRate repository. No convention other than names reflecting the file contents. ● Naming of the document in each deliverable: <ProjectName>_<Document_number>-<Document_Name>_<v.#>
Project and Data Identifiers	Files are stored in the official InteropEHRate repository and data are identified following the initial convention used by FTGM. All other (knowledge and software code) files are identified according to their content.
Community Standard for Metadata Sharing/Integration	Standard and agreed methods will be used to attach metadata to the software code (Javadoc, Maven pom, Git-based versioning).

Automatic Creation of Metadata	No specific automatic creation of metadata is envisioned other than what is usually done by conventional Software development tools.

	DATA ACCESS
Risk Management	Health knowledge data that is publically available and synthetic EHR data that is fake data pose no risks, as no personal or sensitive information is present. Software code that is not (yet) released as open source needs to be protected from open access, following standard software development practice: stored inside a password-protected repository. All software tools and the entire software developing processes of the partner's institution (UNITN) will be formally security-audited by a professional external company. Backup copies of data are created. Furthermore, the parts of software code developed that are not yet fully open source are stored in a GitLab repository that is not accessible from the Internet.
Data Access	The software code, as well as non-personal health knowledge will be accessible on the project-wide GitLab platform (using login + password)
Correct execution of the access process	The project-wide repository is maintained by BYTE, who are responsible for proper access control.
Procedure to address the possibility of a data breach	Data breach with respect to UNITN data is not a risk from the point of view of the InteropEHRate project. Synthetic data and health knowledge are not confidential. The parts of software code that are not open source are only protected with respect to the UNITN IPR and not due to any privacy consideration.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Organization/ Labelling of Data for Easy Identification	In order to be reusable in other projects the software code will be released in a modular form.
Data Sharing Requirements	Apart from software code, in publications UNITN will only use a few examples of non-personal data and, if needed, synthetic EHR data.
Audience for Reuse	System developers interested in adopting the results of the project may reuse software code released as open source.
Restrictions on Re-Use of Data	The license for each software component will be defined by UNITN in accordance with the Consortium Agreement and other partners. Certain software components will be released as open source (licence according to project-wide

	conventions, if they exist); others will remain proprietary under UNITN control.
Publication	The data sets collected/generated in the course of the development of cross-border EHR software will be published on an open website (or the website agreed upon by the entire Consortium) by the last month of the project in accordance with the other InteropEHRate project partners.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	The data sets collected/generated in the course of the development of cross-border EHR software will be archived within a software versioning system with internal access rights.
Data Retention	Data is normally kept for 10-15 years (if not in use anymore), maybe longer depending on the need. The decision is made by the software-developing partner (UNITN) according to internal research policies of this institution.
File Formats	Standard Oracle Java
Data Archives	Open-source code will be archived in common Java repositories, e.g., maven-central.
Long-term Maintenance of Data	UNITN can engage, if necessary, in maintaining the availability of data as long as needed. Beyond maintenance "as is", interventions on the data beyond the project will be up to negotiation based on the long-term needs of project partners and end users.

Table 6 - FAIR Lifecycle of Data Sets: Cross-border electronic health records (EHR)

4.2.2 Smart Electronic Health Records (S-EHR) Mobile App and S-EHR Cloud

4.2.2.1 *Objectives*

As discussed above, the project will develop a e S-EHR mobile app and an optional S-EHR Cloud. An S-EHR is a secure (encrypted) storage system installed on mobile devices that is directly controlled by the citizen. The use of a cloud service is optional. It is different from an EHR, which is under the control of an institution. An S-EHR will allow the citizen to control, and share personal health data with health operators and researchers, in a highly confidential and secure way:

(1) Without the internet: *using the D2D protocol based on short-range connections* between the terminal of the citizen (e.g. a personal smartphone) and the terminal of the health care provider (e.g. a desktop computer in the ambulatory).
(2) On the internet: *The R2D-Access and RDS protocols* will allow use of the S-EHR to exchange health data at distance with healthcare operators and researchers.

4.2.2.2 *Software Development*

A prototype of an S-EHR mobile app is being developed. A prototype of the S-EHR Cloud service will also be developed.

4.2.2.3 *FAIR Lifecycle of Data Sets for Software Coding and Development of S-EHR and S-EHR Cloud Applications*

The following tables provide additional detail regarding the characteristics of the data generated from the tasks associated with developing the S-HER software.

	DATA PRODUCTION AND STORAGE
Data generated/collected by A7	Software Code
Data Format	JavaScript and Kotlin.
Software Tools for creating/processing /visualizing data	Software code can be read by a simple text editor but an IDE (Integrated Development Environment) is highly recommended. In this case, IDE is Android Studio.
Storage and Backup Strategy	GitLab.

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Standards for Documentation and Metadata	No standards currently observed other than those dictated by the programming language.
Best Practices/Guidelines for Data Management	Official Kotlin Guidelines.
Tools for Formatting Data	Android Studio automatically indents the code.
Directory and File Naming Convention	Java naming convention

	DATA ACCESS
Risk Management	The only identified Risk (so far): Stealing of GitLab credentials and access to software code by an unauthorized person. GitLab platform access requires user login and password.

Data Access	Access to software code is provided via a login in GitLab on a developer account. Data processed will not be shared with third parties.
Procedures to Follow a Data Breach	Will be defined in the course of the project.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Data Sharing Requirements	To use an IDE allowing reading and indenting of the code (i.e. Android Studio).
Audience for Reuse	At the moment, only Project partners will use the S-EHR app library. Later (when the project will be on the market) other health application's provider could use this library.
Restrictions on Re-Use of Data	None identified so far.
Publication	As development progresses, the code will be published on GitLab.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	software code will be stored on GitLab
File Formats	Source files on GitLab in Kotlin (.kt)
Data Archives	No data archiving is anticipated.

Table 7 - FAIR Lifecycle of Data Sets: S-EHR Mobile App and S-EHR Cloud

4.2.3 Health Care Professional (HCP) App

4.2.3.1 *Objectives*

InteropEHRate intends to provide hospitals with a secure web app (the Health Care Professional Web App) for health data exchange with the patients. The HCP Web App is based on FHIR standards and will enable HCPs to securely exchange health data of their EHRs with any S-EHR.

4.2.3.2 *Software Development*

A web application used by the healthcare operators to access patients' health records at global level will be prototyped in the project. This app will exploit the D2D and R2D-Emergency protocols for the following purposes:

(1) Import/export data directly from/to the S-HER on the smartphone, including data coming from other healthcare providers;
(2) Import/export data from/to S-EHR cloud for emergency purposes.

4.2.3.3 *FAIR Lifecycle of Data Set for Software Coding and Development of HCP Application*

The following tables further address the data produced in connection with the creation of the Health Care Professionals App.

	DATA PRODUCTION AND STORAGE
Data produced/generated	1. Software Code - necessary for designing and implementing the HCP App solution as well as for the experimental use and testing of the HCP App. New functionalities to the HCP App include the ability to receive images and upload files (e.g. certificates for practitioners and organisations). 2. Randomly generated test data or fake data sets compliant with HL7 FHIR format (fake health data). 3. Deliverables and publications
Data Format	InteropEHRate Deliverable [D2.8] data format and is compliant with HL7 FHIR format.
Reproducibility	The process of data generation is reproducible. The design of HCP App solution ensures the accessibility and reusability of generated data.
Size of the Data	Total size thus far: 1 MB For each category: 100 KB MB for documents and code, GB for videos, KB for test data.
Software tools for creating/processing /visualizing data	HCP App uses WEASIS DICOM viewer (visualizing data).
Use of pre-existing data	The partners will use pre-existing data from coding lists, nomenclatures, etc.
Storage and Backup Strategy	InteropEHRate Partner developing the HCP App will use local databases or specific tapes storage in the partner's institution. The backup and recovery procedures will be implemented in accordance with the requirements and stipulations of ISO 27001.

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Standards for Documentation and Metadata	InteropEHRate partner developing the HCP App will use standardized metadata schemas and encoding schemes as well as persistent and unique identifiers such as DOI (Digital Object Identifier) when implementing HCP App, if applicable for further developments in Year 3.
Tools for Formatting Data	None in this stage of implementing the requirements of HCP App specific for Year 1 and Year 2.
Directory and File Naming Convention	Currently Google Drive (common repository), GitLab (built-in version control, issue tracking, code review etc.), Jira (issue tracking for internal use in the partner's own institution) and Microsoft Project directories are used. No specific naming conventions are used.
Project and Data Identifiers	The responsible partner will use specific coding lists (identifiers assigned for managing the project, e.g. tasks, resources, deliverables,) and standard identification mechanisms compliant with eIDAS Regulation schemes/mechanisms for data [eIDAS].
Community Standard for Metadata Sharing/Integration	The responsible partner will use naming conventions for identifiers specific to Java technology/programming.
Automatic Creation of Metadata	No metadata is automatically created.

	DATA ACCESS
Risk Management	Major risks: data breach, loss of data, security threats, weak authentication (e.g. single factor passwords). Risk Mitigation: The partner responsible for the development of the HCP App (SIMAVI) (certified ISO 27001) has implemented internal procedures in compliance with ISO 27001 (SR ISO/IEC 27001:2013), allowing for access and risk mitigation [ISO 27001]. The developed InteropEHRate protocols include identity management and encryption. Security measures of the HCP-App are based on the API developed by the InteropEHRate project Regular change of passwords and data back-up.
Data Access	Only project partners have access to the data. The procedure for data access is via GitLab tools, username and password is required.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Audience for Reuse	Primarily medical staff will use the HCP App as a demonstrator during pilots.
Restrictions on Re-Use of Data	Too early to define restrictions; these aspects will be set within the analysis stage of the Project.
Publication	Publically available deliverables are published on the InteropEHRate website and submitted to the EC via the EC Portal.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	The data archiving for preservation and long-term access will be implemented in accordance with the requirements and stipulations of ISO 27001 and ISO 9001.
Data Retention	The responsible partner (SIMAVI) will preserve data for at least five years after the end of the project
File Formats	Not yet determined
Data Archives	Open-source code will be archived in common Java repositories
Long-term Maintenance of Data	The Administrator of HCP App and the technical support team of the partner (SIMAVI) will maintain the data for the long-term.

Table 8 - FAIR Lifecycle of Data Sets: HCP App

4.2.4 Interoperability

4.2.4.1 *Objectives*

A key goal of InteropEHRate is to integrate current interoperability infrastructures with new technologies for health data exchange.

The InteropEHRate software specifications will define how to realize within-border and cross-border remote exchange of clinical data and any other health related data between citizens' mobile apps and healthcare providers. Existing standards will be integrated in order to support the coexistence of several levels of interoperability, starting from the exchange of raw data to data adopting a common HL7 FHIR profile and translatable in several natural languages. The selected profiles and terminologies, plus possible extensions, will constitute the InteropEHRate HL7-FHIR profile for EHR interoperability [FHIR].

InteropEHRate will define a set of integrated protocols and conformance criteria for mobile apps, supporting secure and portable local storage and backup, released as open specifications. Moreover, the project will integrate these new protocols with technologies for information extraction and translation, to

reduce the difficulties in health data exchange related to the different terminologies and languages adopted in different European countries and by different healthcare providers.

Two protocols for future interoperability standardization will be designed: one protocol for device to device data exchange (D2D) using Bluetooth short-range communication technology and another protocol for remote data exchange (R2D) that will allow different cross-border EMRs/EHRs to exchange health data of specific patients (the specifications will define the interactions with Health record (HR) indices).

Both protocols will be based on the adoption of a common FHIR profile for the representation of health data and will support different levels of interoperability. An API for D2D and R2D data exchange will also be developed.

The specified protocols and APIs will exploit security protocols and APIs in order to guarantee the protection of any exchanged data.

4.2.4.2 *Software Development*

To develop this software, reusable service components and protocols will be designed and implemented during the Project. This also includes the design of the InteropEHRate Health Services (IHS) and of a prototype of the Health-Record (HRs) Index.

This Health Record Index component will contain only metadata and the design of (public or private) message broker for making the remote exchange of health data between a S-EHR mobile application and research systems (EMRs/EHRs and S-EHR) without cloud storage of HRs more reliable.

The IHS is a set of reusable components, implementing the D2D and R2D for EHR interoperability, interoperable with existing infrastructures for identity management (IID). It also comprises:

(1) The design of a server side component, used by healthcare organizations to share health data contained in their EMR or EHR with the S-EHR and/or to federate their health data with the ones provided by other European organizations,
(2) The design of client side components used by mobile applications (e.g. S-EHR) and web applications (e.g. HCP Web App) for exchanging health data with EHRs and EMRs by means of the D2D protocol and the remote protocol for EHR interoperability.

4.2.4.3 *FAIR Lifecycle of Data Set: Interoperability*

The following tables further represent characteristics of the reusable component and protocol software developed during the Project. Again, FAIR principles are addressed:

	DATA PRODUCTION AND STORAGE
Data generated/collected by UPRC	Software Code
Data Format	Software in Java code (will be updated to other programming languages - if needed.)
Reproducibility	No, this process will not be reproducible. In the case of data loss, backup data

	will be used from the local repositories.
Size of the Data	The current size of data is a few megabytes, and it is generated depending on the implementation stages and on the publication conference calls. This data will change as follows: Software code will change weekly.
Software tools for creating/processing /visualizing data	Microsoft Office Suite, Android Studio, Netbeans, GitLab, Nexus Repository, Google Suite
Use of pre-existing data	No pre-existing data will be used.
Storage and Backup Strategy	The data will be stored locally, and on UPRC personal cloud repositories as a backup strategy. Furthermore, data is stored on external hard disk drives.

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Standards for Documentation and Metadata	No specific standards will be used for the moment.
Tools for Formatting Data	No specific tools needed to format data other than an integrated Development Environment (IDE).
Directory and File Naming Convention	The directory will contain the project's name InteropEHRate, or the acronym IEHR, followed by a descriptive name.
Project and Data Identifiers	The whole project name InteropEHRate will be used for the acronym IEHR, followed by a descriptive name.
Community Standard for Metadata Sharing/Integration	None currently used
Automatic Creation of Metadata	No metadata are automatically created

	DATA ACCESS
Risk Management	The major risk is the possibility of data loss. The responsible partners will keep data in private repositories and will share it with other partners of the project. No formal risk assessment has been conducted for this software. Risk management tactics include updating software regularly, restricting access to the most valuable data and using difficult to decipher passwords.

	Automatic encryption of files when transferring sensitive data, if needed.
Data Access	A unique link will be provided to anyone who wants to gain access.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Audience for Reuse	Data is re-used by project partners for now.
Restrictions on Re-Use of Data	None currently envisioned.
Publication	On the project's website and in GitLab.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	In local repositories and cloud based repositories. (Such as Google Drive and GitLab).
Data Retention	The software code will be retained permanently as it forms the basis for proper functioning of the final mobile app.
File Formats	In their initial file format for all different types of data
Data Archives	No formal archiving.
Long-term Maintenance of Data	The partner (UPRC) that develops the software together with the researchers that participate in the InteropEHRate project will maintain the data for the long-term.

Table 9 - FAIR lifecycle of data sets: Interoperability

4.2.5 Interoperability Profile and Standardisation

A fundamental aspect of a successful integration of already existing and used software and the new InteropEHRate architecture based on an HL7-FHIR profile is the development of a specific customized and standardized InteropEHRate interoperability profile with a high security level.

4.2.5.1 *Objectives and Software Development*

The interoperability HL7-FHIR profile to be adopted by the project is intended to guarantee a high level of syntactic and semantic integration. The HL7-FHIR profile for interoperability will constrain (also with possible extension) the structure of FHIR resources in order to define a single possible representation for each kind of health data to be used by the pilot applications. In addition, requirements regarding the use of value sets are made.

The profile released by the project can be considered a seed for further possible evolutions that could be applied by the stakeholders interested in adopting the platform after the Project.

4.2.5.2 *FAIR Lifecycle of Data Set: Interoperability Profile and Standardization*

The following tables further address the data produced in connection with the creation of the HL7-FHIR interoperability profiles.

	DATA PRODUCTION AND STORAGE
Data Generated/collected by FRAU	1. Human readable documentation of specification of interoperability profile 2. Technical specification of interoperability profile (HL7 FHIR) 3. Software code 4. Specification of privacy and security conformance levels of the S-EHR mobile app 5. Test data/ examples of profile instances.
Data Format	1. technical specification of relevant standards: HL7 Profile, HL7 Resource, Web Service 2. documentation of other relevant standards (e.g. hl7 FHIR international patient summary) and relevant literature for studies / related work: PDF, ZIP 3. software programs, modules or libraries for specification: packages or executables 4. Specification of S-EHR mobile privacy and security conformance levels: WORD, PDF
Reproducibility	All produced data will be subject to a backup strategy, realized by versioning tools (git) and stored in (secure/intranet) cloud services (e.g. Google Drive and GitLab).
Size of the Data	Few MB each, growth rate of the data depends on the provision of additional documentation, specifications or software code.
Use of pre-existing data	• documentation and technical specifications of relevant standards (e.g. HL7 FHIR international patient summary) • relevant literature for studies / related work • software programs, modules or libraries for specification • existing specification / criteria for Specification of S-EHR mobile privacy and security conformance levels (e.g. AppKri-Katalog)
Storage and Backup Strategy	GitLab and Google Drive Repository

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Standards for Documentation and Metadata	The standards used for documentation and metadata align the framework conditions of the standardization bodies and tools. Documentation and specification is realised as an Implementation Guide. We use the tools IG publisher and IG builder. The responsible Partner (FRAU) will use HL7 FHIR as a basis for the specification. FRAU has not decided on the design yet. The profile International Patient Summary will be part of the interoperability profile but there will likely be additional standards and profiles.
Best Practices/Guidelines for Data Management	The partners will create an HL7 FHIR Implementation Guide Resource, which contains the interoperability profile using the tool Forge. The resource will be published with the Implementation Guide Publishing Tool of the HL7 Organization. Draft versions are available in the project Git [HL7]. We will register the Implementation Guide for the Interoperability profile in the HL7 Implementation Guide Registry.
Tools for Formatting Data	In order to check that the data are well formatted the partner (FRAU) uses validation tools for technical specifications.
Directory and File Naming Convention	The directory and file naming convention align with the project standards/project requirements and the framework conditions of the standardization bodies and tools.
Project and Data Identifiers	Project and data identifiers depend on tools / directories that manage the data FRAU will use profile OIDs according to the requirements of standardization bodies. Identifiers in use by FRAU: Identification of the HL7 FHIR Implementation Guide: - Attribute “URL” of the HL7 FHIR Resource Implementation Guide - Datatype “URL” (Uniform Resource Identifier Reference) - Description “Canonical identifier for this implementation guide, represented as a URI (globally unique)”
Community Standard for Metadata Sharing/Integration	FRAU uses the HL7 Community Standard for sharing the Implementation Guide.
Automatic Creation of Metadata	The Publishing Tool of the HL7 Organization creates the metadata.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Re-Use	The data generated/collected by FRAU in the course of the standardization processes will be available and re-usable on websites of standardization bodies (CEN, ISO, HL7): • Documentation of the specification of interoperability profile; • Technical specification of the interoperability profile; • Information on balloting process; • Publications; • Software code; • Software programs & modules; • Test data. FRAU will also make this data available to different research communities, which will be determined in the course of the InteropEHRate project.
Organization/ Labelling of Data for Easy Identification	The responsible partner (FRAU) labels the data according to the requirements of the standardization bodies (such as CEN, ISO, HL7) and of research communities. HL7 FHIR Implementation Guide Resource and the HL7 FHIR software development tools support specific standardized metadata for labelling. Examples for metadata attributes used for easy Identification of Implementation Guides are name, publisher, and version.
Data Sharing Requirements	Research communities usually offer templates and define requirements. Standardization tools support special machine-readable formats. Tools used - IG publisher and IG builder The responsible partner (FRAU) uses the Implementation Guide Publishing Tool for publishing the Implementation Guide of the interoperability profile.
Audience for Reuse	Persons and companies working on standardization in healthcare. Healthcare software vendors. Students of Medical Informatics.
Publication	The publication process will start in September 2019 • On websites of standardization bodies. • In standardization workshops • In scientific conferences.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	The responsible partner (FRAU) does not know yet.
Data Retention	We will register the Implementation Guide for the Interoperability profile in the HL7 Implementation Guide Registry. The profiles will be also stored in GitHub and Simplifier.net
File Formats	PDF Website-Content (depending on the availability of the Web Administration-Tool) HL7 FHIR formats (depending on the versioning and usage of the formats)
Data Archives	Simplifier.net GitHub Artdecor HL7 WIKI pages
Long-term Maintenance of Data	Most likely by Members of standardization bodies

Table 10 - FAIR Lifecycle of Data Sets: Interoperability Profile and Standardisation

4.2.6 Health Record Security and Privacy

4.2.6.1 *Objectives*

The main objective is the proposal of common security protocols and APIs that will be exploited by all the components and applications of the InteropEHRate platform, to guarantee the fulfilment of security and privacy requirements. Another objective is to design and implement reusable software components offering a reference implementation of the designed security protocols and APIs.

4.2.6.2 *Software Development*

A distinct portion of the InteropEHRate protocols concerns the management of security aspects. These sub-protocols and their implementation are managed by the project as distinct results.

The Project will specify security protocols and will develop software for components and the functional primitives regarding identity management, encryption and electronic signing considering the eIDAS Regulation. The responsible partner will develop and implement protocols for encryption mechanics for both:

(1) Health data storage (on mobile devices and cloud services) and

(2) Health data exchange among S-EHR/EHR and Cloud services. Emphasis will be given to the adoption of protocols, regarding hashing, encryption, signing, that are considered battle-tested and unbroken.

All the implemented security libraries per scenario and per component are the following:

Protocol Family	Component	Protocol Libraries	Security Libraries
D2D	S-EHR App	M-D2D-E	M-D2D-SM
	HCP App	T-D2D-E	T-D2D-SM
R2D Access	S-EHR App	M-R2D-E	M-R2D-SM
	Trusted Proxy Server	T-R2D-E	P-R2D-SM
R2D Backup	S-EHR App	R2DWriter	M-R2D-SM
R2D Emergency	HCP App	R2DReader	T-R2D-SM
RDS	S-EHR App	M-RDS	M-RDS-SM
	RRC, CN	S-RDS, IRS	T-RDS-SM

Table 11 - Security libraries per scenario and component

4.2.6.3 *Lifecycle of "FAIR" Data Sets: Health Record Security and Privacy*

The following tables describe additional characteristics of this produced software.

	DATA PRODUCTION AND STORAGE
Data generated/collected by UBITECH	Software code
Size of the Data	Data is software code, which will change very often since it is a work in progress. Currently all implemented libraries are less than 10MB.
Software tools for creating/processing/ visualizing data	For implementation purposes, the partner will use IDE tools to write the software code.
Use of pre-existing data	None
Storage and Backup Strategy	The generated software code will be uploaded for backup in the InteropEHRate GitLab repository.

	Patients' health data is stored in the S-EHR cloud; real data will be stored in the pilot phase. (Further information contained in Section 5 of this document).

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Tools for Formatting Data	For implementation purposes, the responsible partner will use an IDE tool to write the software code and check the data format.
Community Standard for Metadata Sharing/Integration	The partners will use the InteropEHRate private GitLab repository to store and share the software code.
Automatic Creation of Metadata	Configuration files of the software code.

	DATA ACCESS
Risk Management	No disclosure of the source outside of the consortium. Risk of loss: regular backups in the GitLab repository (e.g. every two weeks) Encryption and access control mechanisms
Data Access	Software code should be open source

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Organization/ Labelling of Data for Easy Identification	Not applicable at this stage. Security libraries are unaware of the data, just encrypt, sign and perform security related actions.
Restrictions on Re-Use of Data	No restriction for the consortium members.
Publication	Not applicable at this stage. Security libraries are unaware of the data, just encrypt, sign and perform security related actions.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	Data will be archived in the InteropEHRate GitLab repository and in UBITECH's private GitHub repository.

File Formats	Software code (e.g.: .java files etc.)
Data Archives	InteropEHRate GitLab repository and the responsible partner's (UBITECH) private GitHub repository.
Long-term Maintenance of Data	The responsible partner (UBITECH) will maintain the data in the long term.

Table 12 - FAIR Lifecycle of Data Sets: Health Records Security and Privacy

4.3 Project Pilots

This section of the document will address the data management policy to be implemented during the pilot and validation phase of the project with respect to the processing of personal data. The goal of the Project Pilots is to validate the open specifications and the InteropEHRate Framework by assessing if functional and non-functional requirements are satisfied. This task will involve testing the application with final users, citizens (also referred to as patients). This will be done by experimenting with the application's usage in different international scenarios requiring the cross-border exchange of health data to fulfil the requirements of the three kinds of users – citizens (patients), Health Care professionals and researchers. These tasks are referred to as "pilots".

This experimentation phase will take place at various pilot sites . Personal data will be collected from volunteering patients during the experimentation phase (see tables contained in section 5.3.1, 5.3.2, 5.3.3 and 5.3.4 of this document)

At each pilot site, the InteropEHRate solution will be tested with a small-scale dataset, consistent for healthcare provision of chronic patients, and validated in a real healthcare setting with real patients. In Pilot 1 (Citizen Centred Healthcare Pilot) and Pilot 2 (EHRs Integration for Emergency Pilot), a group of three patients per site will participate in the validation phase. In Pilot 3, (Citizen Centred Medical Research Pilot) 15 patients will participate from each Research Centre. Deliverable D7.1- validation plan [D7.1], gives a comprehensive overview of how final users and other stakeholders will interact with the InteropEHRate applications in the context of each pilot.

The evaluation of the software will consider the following:

- Usability of the developed tools and app.
- Conformance to the requirement specification.
- The level of security privacy and the conformance to the GDPR,
- Advantages brought by availability of a S-EHR connected device allowing:
    - Patients to bring a safe copy of their health data;
    - Healthcare operators to have access, with direct authorization of the patient, to health data;
    - Healthcare operators have access to emergency health data, even in cases where the patient is unresponsive but with the patient's prior consent.
- Clinical accessibility with reference to the possibility of sharing clinical data among different healthcare providers (in different locations) to the completeness of the patient's clinical history (physical exam reports, pathophysiological parameters, etc.).

Possible advantages in the prevention of critical events.

4.3.1 Initial S-EHR Feed (Pilots 1, 2, 3)

The Initial S-EHR feed is the first step, which occurs at all pilot sites, this step is necessary to ensure the proper installation of the S-EHR app by the citizens and the importation of health data before the commencement of testing. This procedure is performed at CHU and FTGM. This initial feed is described in Scenario 0 of deliverable 2.3 [D2.3].

In this process, the volunteering patient installs the S-EHR app and gives his/her consent to store health data on the mobile device, he/she also provides the consent to the Hospital Partner to collect his/her healthcare information from the hospital's EHR system . Thereafter, the patient selects the referral centre that will provide his/her pre-existing health information.

The Hospital's EHR solution implements the interface specified in InteropEHRate project to respond to patient's data requests, selecting the requested information, converting them in the InteropEHRate format and sending the requested information to the patient's phone. Electronic health records of the patients enrolled in the study can be copied from the involved healthcare organisation to a dedicated sandboxed environment (e.g. CHU EHR sandbox). This includes previous encounters, latest vital signs, latest EKG and electro-cardiogram. For example, In the case of CHU, Patient summary is opened by MediSpring, updated and loaded onto the sandbox environment. Finally, Health data of the patient is imported to the S-EHR app from the Electronic Health Records of the selected centre. Imported data is stored on the patients' device in an encrypted format.

4.3.1.1 *Processing Activities*

The following table provides additional detail regarding the processing activities, which take place in the Initial S-EHR Feed.

PROCESSING STEP	DESCRIPTION
Patient enrolment	The volunteering patient reads the information sheet describing the study and signs the informed consent form.
Data preparation	This step may be implemented at pilot sites for the following purposes: 1. Segregate the validation of the InteropEHRate scenarios from its operational activities. 2. Bridge interoperability gaps that cannot be addressed seamlessly by the partners. The actual implementation may vary from one pilot site to another. CHU Liège: Once consent is obtained, electronic health records of the patients enrolled in the study are copied to a dedicated sandboxed environment (CHU EHR sandbox). This includes previous encounters, latest vital signs, latest EKG and electro-cardiogram. Patient summary is opened by MediSpring, updated and loaded onto the sandbox environment. The eIDAS personal identifier of the patient is added as metadata to allow further retrieval.

	FTGM: Once consent is obtained, health information of the patients enrolled in the study are flagged with the respective Anonymization ID and reference ID of the research protocol. Only data that can be requested and transmitted are indexed and referenced to the Pseudo ID and protocol ID. Those data are the subjects of requests and are read and converted only on the patient's request.
Patient onboarding	The patient installs the extended version of the Andaman7 application (Adaman7+) supporting the InteropEHRate protocols. The patient also installs any Android security patches to ensure a secure execution of the Andaman7 application.
Health Data download	Using the Adaman7+ application, the patient selects his/her referral centre, authenticates using an eIDAS compatible mean and downloads his/her health data. Imported data is stored on the patient's device in an encrypted format. The health data imported by the patient was previously converted by the hospital to the common InteropEHRate data format by means of the InteropEHRate data integration platform, configured for the specific hospital.

Table 13 – Initial S-EHR Feed Processing Activities

4.3.1.2 *FAIR Lifecycle of Data Set for Initial S-EHR Feed*

The following tables provide additional details regarding the management of data during the validation of the InteropEHRate systems in the context of the initial S-EHR feed. Please see Deliverable 2.5 [D2.5]for detailed deployment diagrams highlighting different data flows and security measures.

	DATA PRODUCTION AND STORAGE
Data generated/collected	Patient's identification data: name, surname, date of birth, location of birth, gender, country of residence, eIDAS personal identifier. Patients' health data: This study focuses on patients that are followed up for heart failure. The following data shall be present in the patient record: - Main chronic conditions; - Current medications - Patients medical history including reports of past cardio hospitalizations (discharge report, pdf and structured data) (optional); - Patients previous visits (pdf and structured data): diagnosis, treatment plan (next visits, exams, etc., prescribed drugs; - Vital parameters; - Latest instrumental examinations (e.g. EKG, Ultrasounds, Radiology); - Latest bio-humoral values.

	Patients consent forms: paper-based consent.
Data Format	Patient data in HL7 FHIR according to the FHIR profile in the InteropEHRate project. Paper-based consent forms (text) and digital scanned copies.
Reproducibility	None
Size of the Data	10 - 500 Mb
Software Tools for creating/processing /visualizing data	S-EHR App (Andaman 7+): An enhanced version of the Andaman7 application supporting the R2D-Access protocol. The application is installed and used by the patient on his/her mobile phone to collect and exchange his/her health data with HCPs. R2D-Access Service (or simply R2D-Server): A Web server deployed at each pilot site supporting the R2D-Access protocol. Behind the scenes, the server uses the conversion rules from the local hospital data format to the common InteropEHRate data format by means of the InteropEHRate data integration platform. EHR systems of FTGM and CHU: Original data sources InteropEHRate data integration platform: including the InteropEHRate Health Tools (IHT) and InteropEHRate HealthServices (HIS), supports the conversion and translation of structured and unstructured information, and it is configured for the healthcare provider. InteropEHRate Research Services (IRS): components that interoperates with the S-EHR using the RDS protocol, allowing the scientists to engage voluntary citizens at the cross-national level in new research trials and retrospective studies and to receive health data from them. MediSpring®: used by the relevant clinical partner, CHU to validate patient summary in local format. ItsMe: certified eIDAS Qualified Trust Service Provider, used to verify the identity of the Belgian patients.
Storage and Backup Strategy	CHU: Patients consent is stored for a period of 20 years. Admission data & proof of consent are stored in paper form at a secure closet in the cardiology department of CHU. A scanned copy of the patient's consent forms is stored in the CHU IT infrastructure. At the end of the pilot, the CHU EHR sandbox is deleted. FTGM: Patients consent is stored for a period of 10 years in the healthcare facility. Patient's data collected exclusively for the research protocol will be maintained for a period of 7 years.

	HYG: Hygeia is a secondary care hospital. Therefore, patients' consent is stored for a period of at least 20 years along with the medical file. Health data stored on S-EHR: Patients will be asked to delete this data once the pilots are concluded. This may be achieved by uninstalling the app (health data being stored only locally on the device).

	DATA ACCESS
Risk Management	Except the eIDAS infrastructure, all software systems will be deployed in the hospital local network. We rely on the Hospitals' network security systems to avoid access from third parties to patients. Encryption: All data exchanged by the InteropEHRate protocols are transferred with the AES-256 algorithm. The symmetric session key is established every time we need to perform a new exchange with the Diffie-Hellman protocol. Logical Access Control: S-EHR App - password min 8 characters, valid email address (with verification). Requesting a password to access the app is not mandatory but if activated, the password is requested every time the app is launched. Usually, the phone itself is protected by a password as well (in addition to the app). In addition, mobile phones are usually for personal use (reducing as well the risk of illegitimate access). R2D-Server: Accessible only after successful eIDAS authentication. Furthermore, access to the R2DServer is audited. Traceability: S-EHR App - every piece of data entered in the app comes with traceability data (author, timestamp of creation). No modification of data: instead, it is "replaced" with new data (with author and timestamp). However, there is an exception: modification of data without creating a new one is allowed if data has been created less than 3 minutes ago and the same author does the modification. Data is never deleted, It can just be invalidated (the app records a timestamp too for the invalidation action).
Data Access	The patient has only access to the health data related to him/her.

Procedures to address the possibility of a Data Breach	CHU: 1. Breaches of personal/sensitive data are notified to the DPO. 2. The DPO asses the severity of the violation using the ENISA methodology (severity = DPC x EI +CB) 3. If the severity is >= 3, the DPO notifies the competent authority and the people affected by the breach. FTGM: follows the procedure specified in the FTGM document "Data Breach Policy" [FTGM]. HYG: HYG has internal policies and procedures to identify, manage and promptly report any personal/sensitive data breaches in compliance with law and the rights of data subjects. All data breaches are notified to the DPO who, according to the severity, decides whether to further notify the Hellenic Data Protection Authority and the subjects affected by the breach.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Re-Use of Data	Beyond the scope of the pilots, health data and other personal data will not be re-used for purposes not compatible with the initial purpose of the treatment.
Audience for Reuse	Not applicable
Restrictions on Re-Use of Data	Health data and other personal data will not be re-used for purposes not compatible with the initial purpose of the treatment.
Publication	The results of the pilots will be published in scientific publications and disseminations as determined by the project.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	Patients consent forms will be archived (in paper and digital form).

Data Retention	The responsible clinical partner (HYG, FTGM; CHU) will preserve patients consent form for a period of: FTGM: 10 years CHU: 20 years. HYG: at least 20 years
File Formats	Text (paper form originals and digital copies)
Data Archives	CHU: Paper-based consent forms are kept in a secure closet in the cardiology department. FTGM: Paper-based patients' consent are preserved in the secure storage of the healthcare facility. HYG: paper based patients' consent forms are kept in a secure place following certain standards and procedures. Digital copies are archived in clinical partners (HYG, FTGM, SCUBA) IT infrastructure.
Long-term Maintenance of Data	The hospital (HYG, FTGM, CHU) IT department and the local principal investigator. Technical partners (UPRC, A7, SIMAVI, and BYTE UNITN) will not store or maintain patient data beyond the scope of the pilots.

Table 14 - FAIR Lifecycle of Data Set for Initial the S-EHR Feed

4.3.2 Pilot 1: Citizen Centred Healthcare Pilot

4.3.2.1 *Overview*

Pilot 1 seeks to evaluate the exchange of health data between a citizen and a healthcare practitioner (HCP) through the citizens' mobile devices using a 'local' link or device-to-device connection, not involving the use of internet and cloud storage. This phase of the pilot takes place at hospital locations to ensure that the InteropEHRate system is tested in a real environment. The scenario aims to demonstrate the relevant features of the InteropEHRate systems during a real routine patient visit where it is presumed that the citizen is away from their home country and requires a medical visit while abroad.

For this pilot scenario, the platform will be tested in the following countries:

- Greece: ATHENS DIAGNOSTIC AND TREATMENT CENTERS (HYG): 3 patients.
- Italy: GABRIELE MONASTERIO TUSCANY FOUNDATION (FTGM): 3 patients.
- Belgium: UNIVERSITY HOSPITAL CENTER OF LIEGE (CHU): 3 patients

4.3.2.2 *Processing Activities*

The following table provides additional detail regarding the processing activities, which take place in Pilot 1.

Processing step	Description
Bluetooth connection	The patient's smartphone pairs with the HCP terminal for the identification by means of the D2D protocol.
Identity authentication	Using the S-EHR App the patient receives identity data of the Health Organisation and confirms the identity. Thereafter, the Healthcare Professional (HCP) receives the identification data of the patient and confirms the identity. eIDAS Node can be used to support the identification authentication.
Consent Exchange	Through the S-EHR, the patient consents to exchange health data with the admitting organisation. The consent is transmitted to the HCP App and recorded by it for future traceability.
Health Data Exchange	A preconfigured (by the HCP on the HCP App) dataset of patient's data is transferred from the S-EHR App to the HCP App. The HCP App maintains the health data for the duration of the pilot.
Visualisation and language translation	Patient's data is visualized on the HCP App and translated to the language officially related to the Healthcare provider. The conversion and translation is supported by the InteropEHRate data integration platform.
Updating data on the S-EHR	The S-EHR App receives from the HCP App new health data produced in the course of the medical visit, in order to be stored in the patient's smartphone.
Bluetooth connection closure	At the end of the medical visit, the HCP App receives the final message of the connection closure.. Other health data are deleted at the end of the day. For this pilot it is expected that information will be managed only in the HCP App.

Table 15 - Pilot 1 Processing Activities

The complete validation plan is documented in Deliverable D7.1 [D7.1].

4.3.2.3 *FAIR Lifecycle of Data Set for Pilot 1: Citizen Centred Healthcare Pilot*

The following tables provide additional detail regarding the management of data during the validation of the InteropEHRate systems in the context of the Citizen Centred Healthcare Pilot (Pilot 1). Please see

deliverable 2.5 [D2.5]for detailed deployment diagrams highlighting different data flows and security measures.

	DATA PRODUCTION AND STORAGE
Data generated/collected	Patients' identification data: name, surname, date of birth, location of birth, gender, country of residence. Patients' health data: This study focuses on patients that are followed up for heart failure. The following data shall be present in the patient record: - Main chronic conditions; - Current medications - Patients medical history including reports of past cardio hospitalizations (discharge report, pdf and structured data) (optional); - Patients previous visits (pdf and structured data): diagnosis, treatment plan (next visits, exams, etc., prescribed drugs; - Vital parameters; - Latest instrumental examinations (e.g. EKG, Ultrasounds, Radiology); - Latest bio-humoral values. Patients consent forms: paper-based consent and digital consent via the S-EHR app. Answer to questionnaires to assess the InteropEHRate platform. Questionnaires to be completed by participation patients and investigators
Data Format	Patient data in HL7 FHIR according to the FHIR profile in the InteropEHRate project. Paper-based consent forms (text) and digital scanned copies.
Reproducibility	None
Size of the Data	10 - 500 Mb

Software Tools for creating/processing /visualizing data	S-EHR App (Andaman 7+): An enhanced version of the Andaman7 application supporting the D2D protocol. The application is installed and used by the patient on his/her mobile phone to collect and exchange his/her health data with HCPs. The application also manages the patient's consent to process activities. HCP App: A EHR software application, which supports the D2D protocol. The HCP App is used to confirm a patient's identity, transmit patient's data through Bluetooth connection and visualize downloaded patient's data. EHR systems of FTGM and CHU InteropEHRate data integration platform: including the InteropEHRate Health Tools (IHT) and InteropEHRate HealthServices (HIS), supports the conversion and translation of structured and unstructured information, and it is configured for the healthcare provider.
Storage and Backup Strategy	CHU: A scanned copy of the patient's consent forms is stored in the CHU IT infrastructure. Patients consent is stored for a period of 20 years. Admission data & proof of consent are stored in paper form at a secure closet in the cardiology department of CHU. Digital copies are archived in CHU IT infrastructure. Patients' health data is formatted and stored in the hospital's acceptance environment. FTGM: Patients consent is stored for a period of 10 years in the healthcare facility. Data is collected on HCP App, and a copy of the patient's data is delivered to the data owner through the S-EHR App. No other copy of the data is retained in the healthcare facility, except HCP App. At the end of the project the HCP App and its content will be erased. HYG: Hygeia is a secondary care hospital. Therefore, patients' consent is stored for a period of at least 20 years along with the medical file. Health data stored on S-EHR and S-EHR Cloud: Patients will be asked to delete this data once the pilots are concluded. This may be achieved by uninstalling the app (health data being stored only locally on the device) and deleting potential backups the subject might have created.

	DATA ACCESS
Risk Management	Pilot 1 will be deployed in the hospital local network. We rely on the Hospitals' network security systems to avoid access from third parties to patients. Encryption: All data exchanged by the InteropEHRate protocols are transferred with the AES-256 algorithm. The symmetric session key is

	established every time we need to perform a new exchange with the Diffie-Hellman protocol. Anonymization: Answers to questionnaires provided by patients, investigators and healthcare practitioners will be anonymous. Logical Access Control: S-EHR App - password min 8 characters, valid email address (with verification). Requesting a password to access the app is not mandatory but if activated, the password is requested every time the app is launched. Usually, the phone itself is protected by a password as well (in addition to the app). In addition, mobile phones are usually for personal use (reducing as well the risk of illegitimate access). HCP APP - The HCP must have the application installed on his computer. In the HCP App, users have to login in the application and it is necessary to have an active connection to receive data from the S-EHR App. For the transmission of the data, the patient must give consent. Traceability: S-EHR App - every piece of data entered in the app comes with traceability data (author, timestamp of creation). No modification of data: instead, it is "replaced" with new data (with author and timestamp). However, there is an exception: modification of data without creating a new one is allowed if data has been created less than 3 minutes ago and the same author does the modification. Data is never deleted, It can just be invalidated (the app records a timestamp too for the invalidation action). HCP App - The categories of data (e.g. IPS, Laboratory tests) that will be downloaded and viewed by the doctor are audited in the administration section (Audit Information).
Data Access	Health data will be shared with other clinical partners in accordance with the Joint Controllership Agreement. Is not allowed a direct exchange of data among hospitals, but it is allowed for the patient to download on his/her own S-EHR App data coming from different hospitals and then make them available for another hospital, with full control of exchanged data by the data owner. For the validation, local project managers of the hospital will be given access to the patient's personal data (this will be disclosed in the consent forms). The principal investigator will monitor all access to clinical data. No access concerns have been identified. Technical partners whose role in the pilot requires access to pilot data will be granted access. Only staff directly involved in the maintenance of InteropEHRate systems will be given access. The correct execution of the access process will be monitored by the relevant technical partners (UPRC, A7, ENG, SIMAVI, BYTE, and UNITN) and governed by the Data Processing Agreement.
Procedures to address the possibility of a	CHU:

Data Breach	1. Breaches of personal/sensitive data are notified to the DPO. 2. The DPO asses the severity of the violation using the ENISA methodology (severity = DPC x EI +CB) 3. If the severity is >= 3, the DPO notifies the competent authority and the people affected by the breach. FTGM: follows the procedure specified in the FTGM document "Data Breach Policy" [FTGM]. HYG: has in place policies and procedures to identify, manage and promptly report any personal/sensitive data breaches in compliance with law and the rights of data subjects. All data breaches are notified to the DPO who, according to the severity, decides whether to further notify the Hellenic Data Protection Authority and the subjects affected by the breach.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Re-Use of Data	Beyond the scope of the pilots, health data and other personal data will not be re-used for purposes not compatible with the initial purpose of the treatment. Non-personal data (software code and healthcare-related formal knowledge) will be reused by partners or third parties for data that is to be made openly available. Anonymous answers to validation questionnaires will be re-used to publish the results of the pilots and re-used scientific publications and disseminations as determined by the project.
Audience for Reuse	Consortium partners, research institutions, researchers, healthcare practitioners and health application's providers could use the results of the pilots.
Restrictions on Re-Use of Data	Health data and other personal data will not be re-used for purposes not compatible with the initial purpose of the treatment.
Publication	The results of the pilots will be published in scientific publications and disseminations as determined by the project.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	Patients consent forms will be archived (in paper and digital form). Patients' health data needed for further medical care is archived in the hospital's systems, in accordance with national legislation.
Data Retention	The responsible clinical partner (HYG, FTGM; CHU) will preserve patients consent form for a period of: FTGM: 10 years CHU: 20 years. HYG: at least 20 years Data needed for further medical decisions will be recorded and retained in the hospital EHR for: FTGM: 10 years CHU: 30 years. HYG: at least 20 years Anonymous answers to questionnaires are retained.
File Formats	Text (paper form originals and digital copies)
Data Archives	CHU: Paper-based consent forms are kept in a secure closet in the cardiology department. FTGM: Paper-based patients' consent are preserved in the secure storage of the healthcare facility. HYG: paper based patients' consent forms are kept in a secure place following certain standards and procedures. Digital copies are archived in clinical partners (HYG, FTGM, SCUBA) IT infrastructure.
Long-term Maintenance of Data	The hospital (HYG, FTGM, CHU) IT department and the local principal investigator. Technical partners (UPRC, A7, SIMAVI, BYTE and UNITN) will not store or maintain patient data beyond the scope of the pilots.

Table 16 - FAIR Lifecycle of Data Set for Pilot 1: Citizen Centred Healthcare Pilot

4.3.3 Pilot 2: Electronic Health Records (EHRs) Integration for Emergency Pilot

4.3.3.1 *Overview*

Pilot 2 seeks to evaluate the R2D Emergency protocol, which supports the exchange of encrypted health data between authorised health care practitioners through the HCP App and a compatible S-EHR Cloud during a medical emergency. Similar to Pilot 1, this phase of the pilot takes place at hospital locations to ensure that the InteropEHRate system is tested in a real environment during an emergency. The purpose of this scenario is to show how HCPs may access and contribute to Patients' health data when a S-EHR is not available or when the Patient cannot use it, in particular in an emergency situation utilising the S-EHR Cloud.

For this scenario, the InteropEHRate protocol and applications will be tested in the following countries:

- Romania: BAGDASAR-ARSENI EMERGENCY CLINICAL HOSPITAL (SCUBA): 3 patients.
- Italy: GABRIELE MONASTERIO TUSCANY FOUNDATION (FTGM): 3 patients.
- Belgium: UNIVERSITY HOSPITAL CENTER OF LIEGE (CHU): 3 patients.

4.3.3.2 *Processing Activities*

The following table provides additional detail regarding the processing activities, which take place in Pilot 2.

PROCESSING STEP	DESCRIPTION
Consent Exchange	The Patient gives informed consent for the storage and management of an encrypted copy of his/her personal health data and identity data on S-EHR Cloud. The Patient consents in the S-EHR app to the emergency identification by means of an emergency identity token. Furthermore, the patient consents to share his or her health data stored on the S-EHR Cloud with HCPs in an emergency.
Authentication of Identity	The patient's emergency identity token represented by a QR-code is used in emergencies by authorized HCPs to verify the patient's identity, access and decrypt the data contained in S-EHR Cloud. Each HCP involved in this scenario has a digital identity issued by a national or local authority or healthcare provider recognised by the hospital and sent to the S-EHR Cloud infrastructure, and associated in a trusted way to his/her qualification.
Health Data Exchange	HCP App allows the HCPs involved in the patient's treatment to access the S-EHR Cloud that will authorize them to access the (emergency) health data.
Visualisation and language translation	Once the HCP App reads the QR Code, the health data is downloaded from the S-EHR Cloud, decrypted and visualized by HCPs currently involved in the patient's treatment process. The InteropEHRate data integration platform also translates

	patients' data into the HCPs natural language.
Updating data on the S-EHR	At the end of the hospitalisation, the HCP writes the discharge summary for the patient on the HCP app, and the S-EHR Cloud is updated with the produced Discharge Summary.

Table 17 - Pilot 2 Processing Activities

The complete validation plan is documented in Deliverable 7.1 [D.7.1].

4.3.3.3 *FAIR Lifecycle of Data Set for Pilot 2: EHRs Integration for Emergency Pilot*

The following tables provide additional detail regarding the management of data during the validation of the InteropEHRate systems in the context of remote to device (R2D), Emergency Access (scenario 2).

	DATA PRODUCTION AND STORAGE
Data generated/collected	1. Patients' identification data: name, surname, date of birth, location of birth, gender, country of residence. 2. Patients' Biometric data: image of the patients' face for identification purposes. 3. Patients' Summary (for emergency) 4. Patients' health data: • Allergies , intolerance • Main Chronic Conditions ○ Ischemic heart disease ○ Heart failure ○ Pulmonary disease ○ Abnormal kidney function ○ Abnormal liver function ○ Previous major surgery ○ Active malignancy • Current Medications [previous medications] • Patients' medical history: Reports of past cardio hospitalizations (discharge report, pdf and structured data) (optional). Previous visits (pdf and structured data), including diagnosis, treatment plan (next visits, exams, etc.) and prescribed drugs; • Vital parameters for the last ambulatory visit; • Latest instrumental examinations (e.g. EKG, Ultrasounds, Radiology, angiography); • Latest bio-humoral values; 5. Patients consent: paper-based consent for participation in the study and digital consent to processing activities via the S-EHR app and S-EHR Cloud.

	6. Answer to questionnaires to assess the InteropEHRate platform. Answers provided by participating patients, investigators and healthcare practitioners.
Data Format	Patient summary and health-related data in HL7 FHIR according to the FHIR profile in the InteropEHRate project. Paper-based consent forms (text) and digital scanned copies.
Reproducibility	None
Size of the Data	10-500 Mb, depending on the data available on S-EHR Cloud
Software Tools for creating/processing /visualizing data	S-EHR App (Andaman 7+): An enhanced version of the Andaman7 application supporting the R2D protocol. The patient on his /her mobile phone installs the application. In this Pilot, the App is used as a data source to store information on the S-EHR Cloud and generates the Emergency Tag (QR Code). Emergency Identity token: generated by the S-EHR app and used to identify the patient in an emergency. S-EHR Cloud: contents of the patient's S-EHR are replicated and stored on the S-EHR Cloud. Identity data, including a photo of the face of the patient, are uploaded in the S-EHR cloud, to identify the patient in the emergency scenario. HCP App: A EHR software application, which supports the R2D protocol. The HCP App is used to confirm a patient's identity, transmit patient's data from the S-EHR Cloud and visualize patient's data. The InteropEHRate data integration platform: supports the conversion and translation of structured and unstructured information, and it is configured for the healthcare provider.
Storage and Backup Strategy	FTGM: Patients consent is stored for a period of 10 years in the healthcare facility. Data is collected on HCP App, and a copy of the patient's data is delivered to the data owner through the S-EHR app. No other copy of the data is retained in the healthcare facility, except HCP App. At the end of the project the HCP App and its content will be erased. CHU: A scanned copy of the patient's consent forms is stored in the CHU IT infrastructure. Patients consent is stored for a period of 20 years. Admission data & proof of consent are stored in paper form at a secure closet in the cardiology department of CHU. Digital copies are archived in CHU IT infrastructure. Patients'

	health data is formatted and stored in the hospital's acceptance environment. SCUBA: paper consent forms will be stored under lock in the research lab for 30 years. Health data stored on S-EHR and S-EHR Cloud: Patients will be asked to delete this data once the pilots are concluded. This may be achieved by uninstalling the app (health data being stored only locally on the device) and deleting potential backups the subject might have created.

	DATA ACCESS
Risk Management	Health data repositories involved in pilots 1 and 2 will be deployed on the hospital's local network, thus restricting the access of third parties to patient data. Pilot 2 will additionally exploit the S-EHR Cloud repository deployed on the infrastructure provided by BYTE. Encryption: All data exchanged by the InteropEHRate protocols are transferred with the AES-256 algorithm. For data storage in S-EHR Cloud, the data is stored encrypted again with AES-256. Health data decryption is performed only by specific operations executed by the HCP App and the S-EHR App. No decryption of health data is performed by any operation executed by S-EHR Cloud. Anonymization: Answers to questionnaires provided by patients, investigators and healthcare practitioners will be anonymous. Partitioning Data: On the S-EHR Cloud, all data is stored as encrypted objects under the MinIO server. Each citizen is assigned to a different bucket or buckets. Personal information of the citizen is stored separately on MongoDB, which by default splits the data into chunks. S-EHR Cloud: In the S-EHR Cloud, users are registered and logged in through username/password. For emergencies, the QR-code provided by the Citizen is scanned by the HCP. The QR-code contains information related to the S-EHR Cloud the citizen used in order to back up their health data. Based on this information, the HCP requests access for that HO. Only upon the provision of the appropriate attributes that can be used to authorize the HO from the ABAC engine, the HO and its HCPs are granted access to the S-EHR Cloud. Every data exchange or data manipulation is audited by the S-EHR Cloud service. In addition, auditing information regarding the list of the HCPs that were granted access to the S-EHR Cloud is also kept. The auditing information can be made available to the citizen's whose health data is stored in the S-EHR Cloud. HCP App: The HCP must have the application installed on his computer. In the HCP App, users have to login in the application and must have an active internet connection to receive data from the S-EHR Cloud and scan the QR code for

	access to the S-EHR Cloud. For the transmission of data, the patient must give his consent. The categories of data (e.g. IPS, Laboratory tests) that will be downloaded and viewed by the doctor are audited in the administration section (Audit Information).
Data Access	Health data will be shared with other clinical partners in accordance with the Joint Controllership Agreement. Direct exchange of data among hospitals is not permitted. However, the data subject is allowed to download his/her own S-EHR App data coming from different hospitals and then make them available for another hospital, with full control of exchanged data by the data subject. For the validation, local project managers of the hospital will be given access to the patient's personal data (this will be disclosed in the consent forms). The principal investigator will monitor all access to clinical data. No access concerns have been identified. Technical partners whose role in the pilot requires access to pilot data will be granted access. Only staff directly involved in the maintenance of InteropEHRate systems will be given access. The correct execution of the access process will be monitored by the relevant technical partners (UPRC, A7, ENG, SIMAVI, BYTE UNITN) as governed by the Data Processing Agreement.
Procedures to address the possibility of a Data Breach	CHU: 1. Breaches of personal/sensitive data are notified to the DPO. 2. The DPO asses the severity of the violation using the ENISA methodology (severity = DPC x EI +CB) 3. If the severity is >= 3, the DPO notifies the competent authority and the people affected by the breach. FTGM: follows the procedure specified in the FTGM document “Data Breach Policy” [FTGM]. SCUBA: Data breaches are reported to the DPO, which evaluates the severity and notifies the authorities and the affected people.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Re-Use of Data	Beyond the scope of the pilots, clinical/patient data and other personal data will not be re-used for purposes not compatible with the initial purpose of the treatment. Anonymous answers to validation questionnaires will be re-used. The results of the pilots will be published in scientific publications and disseminations as determined by the project.
Audience for Reuse	Consortium partners, research institutions, researchers, healthcare practitioners and health application's providers could use the results of the pilots.
Restrictions on Re-Use of Data	Health data and other personal data will not be re-used for purposes not compatible with the initial purpose of the treatment.
Publication	The results of the pilots will be published in scientific publications and disseminations as determined by the project.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	Patients consent forms will be archived (in paper and digital form). SCUBA - paper-based consent forms will be stored under lock in the research lab. Patients' health data needed for further medical care is archived in the hospital's systems, in accordance with national legislation.
Data Retention	The responsible clinical partners (SCUBA, FTGM, CHU) will preserve patients' consent form for: FTGM: 10 years CHU: 20 years. SCUBA: 30 years. Data needed for further medical decisions will be recorded and retained in the hospital EHR for: FTGM: no limit.

	CHU: 30 years. SCUBA: 30 years. Anonymous answers to questionnaires are retained.
File Formats	Text (paper form originals and digital copies)
Data Archives	Paper-based consent forms are archived and digital copies are archived in the IT infrastructure of the relevant clinical partners (SCUBA, FTGM and CHU).
Long-term Maintenance of Data	The hospital (SCUBA, FTGM and CHU) and the local principal investigator are responsible. Technical partners (UPRC, A7, SIMAVI, BYTE and UNITN) will not store or maintain patient data beyond the scope of the pilots.

Table 18- FAIR Lifecycle of Data Set for Pilot 2: EHRs integration for Emergency Pilot

4.3.4 Pilot 3: Citizen Centred Medical Research Pilots

4.3.4.1 *Overview*

This pilot is intended to validate the InteropEHRate Research scenario. The Research scenario is intended to show how a patient's clinical data can be shared through the InteropEHRate Open Research Network with research institutions for research purposes. The InteropEHRate Open Research Network allows participating researchers to enrol citizens in their research studies (described by specific research protocols) and collect health data for the studies directly from the enrolled citizens. Researchers belonging to the InteropEHRate Open Research Network share a common vocabulary, defined by the InteropEHRate profiles, used to refer to any health data required by the research studies performed on the InteropEHRate Open Research Network.
For this scenario, the platform will be tested in the following countries:

- Italy: GABRIELE MONASTERIO FOUNDATION (FTGM): 15 patients.
- Belgium: UNIVERSITY HOSPITAL OF LIEGE (CHU): 15 patients.

4.3.4.2 *Processing Activities*

The following table provides additional detail regarding the processing activities, which take place in Pilot 3.

PROCESSING STEPS	DESCRIPTION
Publishing of a specific research study	Using the Central Node, A Research Centre publishes an approved Research Data Description (RDD), describing the research studies that the citizen may participate in.
Patients enrolment in research study	S-EHR app, using the RDS protocol queries the Central Node to retrieve published RDDs, which are open in the current period. If the citizen satisfies the enrolment criteria of the RDD, S-EHR app provides details about the research study and data collection. The citizen is requested to consent to enrolment and data collection.
Health Data Exchange	The S-EHR app transmits pseudonymized and anonymised health data to the Reference Research Centre.
Withdrawal from the research study	The mobile app, using the RDS protocol sends to the Research Centre Information System the notification that the Citizen will not participate in a specific research study anymore.

Table 19- Pilot 3: Processing Activities

4.3.4.3 *FAIR Lifecycle of Data Set for Pilot 3: Citizen Centred Medical Research Pilot*

The following tables provide additional detail regarding the management of data during the validation of the InteropEHRate systems in the context of the Citizen Centred Medical Research Pilot.

	DATA PRODUCTION AND STORAGE
Data generated/collected	Patients' identification data: name, surname, date of birth, location of birth, gender, country of residence. Associated pseudonym. Patients health data: ● Allergies ● Year of hypertension diagnosis ● blood pressure measurement SYS/DIA (mmHg/mmHg) ● Latest creatinine (mg/dL) ● Current Medications ● EKG report and signal ● Echocardiogram report and video ○ Latest left ventricular ejection fraction (%) ○ Latest interventricular septum thickness (mm) ● Type of symptoms for adverse side effects

	Patients consent: paper-based consent forms for participation in the study and digital consent to processing activities via the S-EHR app. Answers to questionnaires to assess the InteropEHRate platform. Answers provided by participating patients and investigators.
Data Format	Patient data described by the clinical research protocol formatted according to InteropEHRate HL7 FHIR protocol. Paper-based consent forms (text) and digital scanned copies.
Reproducibility	None
Size of the Data	10-500 Mb
Software Tools for creating/processing /visualizing data	S-EHR App (Andaman 7+): An enhanced version of the Andaman7 application supporting the exchange of data for research purposes, through the InteropEHRate Open Research Network. The patient on his /her mobile phone installs the application. The S-EHR receives formal research study descriptions from the InteropEHRate Open Research Network and, if the patient is eligible and willing to participate in a given study, anonymously shares his/her health data for the purposes of research. InteropEHRate Research Services: software components that, when deployed within hospitals, universities, or research centres constituting a Research Network, allow the management of automated data collection from patients' mobile devices for the purposes of multi-centric research studies. The InteropEHRate Research Services allow (1) the publication of formal research study descriptions destined to mobile devices connected to the Research Network; (2) the management of patient consent to participation in specific studies; as well as (3) the collection of anonymised health data retrieved from the mobile device.
Storage and Backup Strategy	Patients consent (to participate in the clinical study) is stored by the local principal investigator on paper form at the hospital site. Digital copies may be archived in hospitals IT infrastructure. Patients consent is stored for a period of 10 years. Patients' health data is stored in the pilots' research facility for a period of 7 years.

	DATA ACCESS
Risk Management	Health data of pilot 3 will be deployed on the citizen's HCP app and on the research facilities' local network. We rely on the research facilities' network security systems to prevent access from third unauthorized parties. Encryption: All data exchanged by the InteropEHRate protocols are transferred with the AES-256 algorithm. The symmetric session key is established every time we need to perform a new exchange with the Diffie-Hellman protocol. Anonymization & Pseudonymisation: Data anonymization and pseudonymisation will both be implemented in the platform, and pseudonymisation will be used in the Research Scenario. More specifically, the citizens' data will either be anonymized or pseudonymized based on the research design description. In data anonymization all data, which is not needed for the study, will be deleted, whereas in data pseudonymisation all data that is not needed for the study, will be replaced with either a pseudo-identity or a pseudonym. The pseudo-identities will be generated at the Reference Research Centre, whereas the Pseudonym Provider (Trusted Third Party) will create the pseudonyms. Data pseudonymisation will be implemented in case that the data controller should reverse the process and be led to the identity of the citizen (e.g. for emergency and public health purposes). Otherwise, data anonymization will be implemented, which is an irreversible process. Answers to questionnaires provided by patients and investigators at the end of the study will be anonymous. Logical Access Control: S-EHR App - password min 8 characters, valid email address (with verification). Requesting a password to access the app is not mandatory but if activated, the password is requested every time the app is launched. Usually, the phone itself is protected by a password as well (in addition to the app). In addition, mobile phones are usually for personal use (reducing as well the risk of illegitimate access). Traceability: S-EHR App - every piece of data entered in the app comes with traceability data (author, timestamp of creation). No modification of data: instead, it is "replaced" with new data (with author and timestamp). N.B. There is an exception: modification of data without creating a new one is allowed if data has been created less than 3 minutes ago and is from the same author. Data is never deleted. It can just be invalidated (the app records a timestamp too for the invalidation action).
Data Access	Clinical data will be shared with other clinical partners in accordance with the Joint Controllership Agreement. For the validation, local project managers of the hospital will be given access to the patient's personal data (this will be

	disclosed in the consent forms). The principal investigator will monitor all access to clinical data. No access concerns have been identified. Technical partners whose role in the pilot requires access to pilot data will be granted access. Only staff directly involved in the maintenance of InteropEHRate systems will be given access. The correct execution of the access process will be monitored by the relevant technical partners (UPRC, A7, ENG, SIMAVI, BYTE and UNITN) as governed by the Data Processing Agreement.
Procedures to address the possibility of a Data Breach	CHU: 1. Breaches of personal/sensitive data are notified to the DPO. 2. The DPO asses the severity of the violation using the ENISA methodology (severity = DPC x EI +CB) 3. If the severity is >= 3, the DPO notifies the competent authority and the people affected by the breach. FTGM: follows the procedure specified in the FTGM document "Data Breach Policy".

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Re-Use of Data	Beyond the scope of the pilots, clinical/patient data and other personal data will not be re-used for purposes not compatible with the initial purpose of the treatment Non-personal data (software code and healthcare-related formal knowledge) will be reused by partners or third parties for data that is to be made openly available. Anonymous answers to validation questionnaires will be re-used to publish the results of the pilots and re-used scientific publications and disseminations as determined by the project.
Audience for Reuse	Consortium partners, research institutions, researchers, healthcare practitioners and health application's providers could use the results of the pilots.
Restrictions on Re-Use of Data	Health data and other personal data will not be re-used for purposes not

	compatible with the initial purpose of the treatment.
Publication	The results of the pilots will be published in scientific publications and disseminations as determined by the project.

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	Patients' consent forms will be archived (in paper and digital form).
Data Retention	The responsible research facilities (FTGM and CHU) will preserve patients' consent form for a period of 10 years. Anonymous answers to questionnaires are retained.
File Formats	Text (paper form originals and digital copies)
Data Archives	Paper-based consent forms are archived and digital copies are archived in the IT infrastructure of the relevant clinical partners (FTGM, CHU).
Long-term Maintenance of Data	The research facilities (FTGM, CHU) and the local principal investigator are responsible. Technical partners (UPRC, A7, SIMAVI, BYTE and UNITN) will not store or maintain patient data beyond the scope of the pilots.

Table 20 - FAIR Lifecycle of Data Sets for Pilot 3: Citizen Centred Medical Research Pilot

4.4 Data Anonymization Approach

This section of the document briefly outlines the approach to anonymization, which will be implemented in the InteropEHRate Pilots.

Data anonymization is the process by which the personally identifiable information is deleted or modified, in order not to be connected to an individual anymore [PIWIK PRO]. After data anonymization is implemented, it is impossible for someone to identify an individual in a dataset [PIWIK PRO].

The personally identifiable information can be divided into:

- Direct identifiers (Unique identifiers): An attribute that is unique and can be linked to a specific individual, e.g., name, surname, social security number.
- Indirect identifiers (Quasi-identifiers): A group of attributes that can be linked to specific individuals, e.g., zip code, age, gender.
- Sensitive data: Attributes which must be confidential, since they may trespass the individuals' rights in case that they are exposed. e.g., health data.

A widespread method of anonymization is k-anonymity. K-anonymity can be implemented on both structured and unstructured data [MEHTA], [MOTWANI]. In this technique, the indirect identifiers are modified in a way that a specific person's data cannot be distinguished from at least k – 1 persons in the same group [MOTWANI], [TERROVITIS]. This can be achieved by suppression (hiding data) and/or generalization (replacement of attributes with broader categories) [MOTWANI], [TERROVITIS]. As a result, no one can determine which data corresponds to a specific individual. In addition, the probability of redefining a record of the dataset is p> 1/k, where k are the records of each group where the values are the same.

However, in case of unstructured data, this method may be very complex, since it requires the classification of the personally identifiable information in the above-mentioned categories [MEHTA]. A way to address this issue is to convert the unstructured data to structured data by applying Natural Language Processing (NLP) methods to the data, and then use the k-anonymity method to these structured data [MEHTA].

In the context of InteropEHRate, the unstructured data will not be anonymized on the citizen's phone. More specifically, there will be two versions of the citizen's health data, the original version of his/her data and an anonymized version, which will not include any information that can lead to the identification of the citizen [D4.8], [D4.10], [D6.8]. The anonymized data will be uploaded to the S-EHR Application along with the original version of the citizen's data, and each time a research study is conducted, the citizen's anonymized dataset will be sent to the Reference Research Centre [D4.8], [D4.10], [D6.8]. The reason why pre-anonymized data will be uploaded on the citizen's phone is because it is computationally difficult to perform anonymization operations on the mobile phone [D6.8]. The anonymization of exchanged health data is needed only in the Research Pilot, and in particular the anonymization of unstructured data will be implemented by the data providers of the pilot site by means of any tools already used by that Pilot site. However, there are many tools in order to perform data anonymization and generate pre-anonymized data that will be used for research studies [ARX.DEIDENTIFIER.ORG].

The anonymization approach is further detailed in deliverables D7.1 and D6.8, taking into consideration the rights and freedoms of data subjects.

5 DATA PROTECTION AND ETHICAL ASPECTS

Data protection is a central issue for research ethics. A fundamental human right, enshrined in the EU Charter of Fundamental Rights, provides all individuals with control over the way information about them is collected and used.[4] 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) grants everyone the right to the protection of personal data concerning him or her and GDPR lays down rules relating to the protection of natural persons with regard to the processing of personal data.

> *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).*

Whenever personal data - information relating to an identified or identifiable natural person is processed[5] special care has to be taken. In research settings, data protection imposes obligations on researchers to provide research subjects with detailed information about what will happen to the personal data that they collect. 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 explicit consent [Article 9(2)(a) GDPR].

In the InteropEHRate project, all data processing will comply with EU law as well as national data laws and will follow the guidelines on "Ethics and Data Protection"[EC]. It will ensure that any partners, contractors or service providers that process research data at the InteropEHRate 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.

5.1 The Ethical Package

In order to include volunteers in the study several measures have to be taken in advance. In the InteropEHRate project four tools, collectively referred to as the Ethical Package will be used to ensure full compliance with data protection and ethical standards: Declaration of Compliance, Data Protection Impact Assessment, Data sharing agreement(s) and Informed Consent. These documents forming the Ethical Package will be checked and approved by ethical committees at each pilot site and have been validated by the InteropEHRate Consortium

The following sections provide more information on the documents that form part of the Ethical Package.

[4] Article 8 EU Charter of Fundamental Rights.

[5] 'Processing of personal data' means any operation (or set of operations) performed on personal data, either manually or by automatic means such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction (Art 4(2) GDPR)

5.1.1 Declaration of Compliance

All partners who will have access to personal data of patients will sign a declaration of compliance (ANNEX 2). This is a legal document in which partners proclaim that they have and will continue to comply with all legal and ethical requirements when processing personal data. In some circumstances, Member States are able to introduce exemptions and derogations to the provisions of the General Data Protection Regulation (GDPR). This Declaration of Compliance states all GDPR derogations to the processing of special categories of personal data and processing for scientific research purposes.

5.1.2 Data Protection Impact Assessment

Given the sensitive nature of health data, processing of such data poses a risk to the rights and freedoms of users. Article 35 GDPR requires controllers to carry out a Data Protection Impact Assessment (DPIA) in such situations in order to identify and minimize data protection risks in processing. The DPIA aims to achieve the following objectives:

- Describe the nature, scope, context and purpose of the processing;
- Assess necessity, proportionality and compliance measures;
- Identify and assess risks to individuals;
- Identify any additional measures to mitigate risks.

The InteropEHRate partners have performed an analysis for each scenario using software provided by the French Data Protection Authority (CNIL). The Ethical Package consists of two DPIAs: one of which is focused on Pilot 1 (D2D Health Record Exchange) and Pilot 2 (R2D Emergency Access) and a separate DPIA focusing on Pilot 3 (Research Scenario). This distinction is made because Pilot 1 and 2 take place in the context of real medical visits at hospital pilot sites. On the other hand, Pilot 3 takes place in the context of research for purposes limited to research.

5.1.3 Data Sharing Agreements

All partners who jointly determine the purpose and means of the processing operation will be considered as joint controllers [EDPB]. Joint controllers will sign a written joint controllership agreement prior to the processing of personal data and the commencement of the pilots. The joint controllership agreement will be drafted by UNIVIE and signed by the following partners as joint controllers, FTGM, SCUBA, CHU, and HYG. In addition, a Data Processing Agreement will be concluded between all data controllers and data processors. The following partners are data processors in the context of the pilots: A7, ENG, BYTE, SIMAVI, UBITECH, UPRC and UNITN. These contracts will determine the respective responsibilities of each partner and ensure compliance with the obligations under the GDPR. The essence of the agreements will be made available to the data subject as required by Article 26(2) GDPR.

Data Controllers: SCUBA, FTGM, CHU, HYG
Data Processors: A7, ENG, BYTE, SIMAVI, UBITECH, UPRC, UNITN
Hosts of the data and services: SCUBA, FTGM, CHU, HYG, BYTE, UBITECH, UNITN
Operating the algorithms and interactivity: A7, ENG, BYTE, SIMAVI, UBITECH, UPRC, UNITN

5.1.4 Informed Consent

The validation phase is premised on the processing of health data. Under Article 9 of the GDPR, personal data concerning health is classified under the special categories of personal data; the processing of such data is prohibited. However, this prohibition is subject to some exceptions outlined in Article 9(2) GDPR.

In Accordance with Article 9(2) (a) health data may be processed if

> *The data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provides that the prohibition referred to in paragraph 1 may not be lifted by the data subject.*

In the context of InteropEHRate, consent forms the legal basis for the initial participation in the pilot study as well as all processing activities, which take place in each pilot. Article 4 GDPR defines consent as a

> *Freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.*

According to WP29 Guidelines on consent, the term 'explicit' implies that data subjects must give an express statement of consent [WP 29]. A written statement, including electronic means (Recital 32 GDPR), can fulfil this requirement.

For the validation phase, partners will seek the consent of participants in two different ways:

1. Consent to participate in the validation and testing will be sought through paper-based consent forms with corresponding information sheets. Patients will be required to sign the consent form prior to participating in the pilots.
2. Consent to specific processing activities, which take place on the S-EHR app, HCP app and S-EHR cloud, will be sought electronically via the S-EHR app.

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 [Article 7(2) GDPR].

Once the testing has commenced in each pilot, the consent of the participants will be sought for each processing activity. Consent is required at various stages including when installing the S-EHR applications, exchanging data, cloud storage and for participation in research data sharing. In each instance, the user will be asked to actively consent by ticking a box on his or her smart device and will be given the opportunity to refuse or withdraw his or her consent at any time. This gives user control over data processing and adheres to GDPR principles of lawfulness, fairness and transparency [Article 5(1)(a), GDPR]. The process of obtaining informed consent on the S-EHR is detailed in deliverable 6.10 of the InteropEHRate project [D6.10].

6 CONCLUSION

This document describes the data management policy adopted in the InteropEHRate project as well as in the upcoming pilot and validation phase. The DMP is a living document, which is updated whenever significant changes arise in the project.

7 REFERENCES

[ARX.DEIDENTIFIER.ORG] Arx.deidentifier.org Website: https://arx.deidentifier.org/overview/related-software/

[CEN] European Committee for Standardisation (CEN). https://www.cen.eu/Pages/default.aspx.

[D1.5] InteropEHRate Consortium, D1.5-Quality Plan, 2019.

[D1.7] InteropEHRate Consortium, D1.7-Data Management Plan – V1, 2019.www.interopehrate.eu/resources.

[D2.3] InteropEHRate Consortium. D2.3-User Requirements for Cross-Border HR Integration-V3, 2021. www.interopehrate.eu/resources.

[D2.5] InteropEHRate Consortium. D2.5. InteropEHRate Project. www.interopehrate.eu/resources.

[D2.8] InteropEHRate Consortium, D2.8- FHIR Profile for EHR Interoperability.

[D2.10] InteropEHRate Consortium. D2.10 -Development and Testing Environment., 2019.www.interopehrate.eu/resources.

[D4.8] InteropEHRate Consortium, D4.8-Specification of protocol and APIs for research health data sharing - V1, 2021. www.interopehrate.eu/resources

[D4.10] InteropEHRate Consortium, D4.10- Design of the libraries for health data sharing for research - V1, 2021. www.interopehrate.eu/resources

[D6.8] InteropEHRate Consortium, D6.8-Design of a mobile service for data anonymization and aggregation, 2021. www.interopehrate.eu/resources

[D6.10] InteropEHRate Consortium, D6.10- S-EHR Mobile App - V2, 2021 www.interopehrate.eu/resources.

[D7.1] InteropEHRate Consortium, D7.1- Experimentation Scenarios and Validation Plan, 2021. www.interopehrate.eu/resources.

[DAMA] DAMA, United Kingdom. The Six Primary Dimensions for Data Quality Assessment, Defining Data Quality Dimensions. (2013). https://damauk.wildapricot.org/resources/Documents/DAMA%20UK%20DQ%20Dimensions%20White%20Paper2020.pdf.

[DATA ARCHIVE] UK Data Archive, University of Essex. Data Life Cycle & Data Management Planning. (2013). https://ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf.

[eHDSI] eHealth Digital Service Infrastructure. EHealth DSI Operations

[eIDAS] REGULATION (EU) No 910/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC. https://ec.europa.eu/futurium/en/system/files/ged/eidas_regulation.pdf.

[EC] European Commission, Exchange of Electronic Health Records across the EU https://digital-strategy.ec.europa.eu/en/policies/electronic-health-records.

[EC] European Commission, Ethics and Data Protection. (November 2018). https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/ethics/h2020_hi_ethics-data-protection_en.pdf.

[ERC] European Research Council (ERC), Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020. (2017). https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf.

[EDPB] European Data Protection Board. Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR. https://edpb.europa.eu/sites/default/files/consultation/edpb_guidelines_202007_controllerprocessor_en.pdf.

[EU] European Union. Charter of Fundamental Rights of the European Union (2012/C 326/02). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN.

[FAIR] European Commission Directorate-General for Research and Innovation. Guidelines on FAIR Data Management in Horizon 2020. (2016). https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[FHIR] Fast Healthcare Interoperability Resources (FHIR) is a standard for health care data exchange, published by HL7®. https://www.hl7.org/fhir/overview.html.

[FTGM] Fondazione Monasterio la ricerca che cura, Data Breach Policy. https://www.monasterio.it/.

[GDPR] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

[GOOGLE] Google Cloud Whitepaper. Google Workspace Security Whitepaper. (October 2020). https://storage.googleapis.com/gfw-touched-accounts-pdfs/google-cloud-security-and-compliance-whitepaper.pdf.

[HL7] Health Level Seven International. http://www.hl7.org/.

[ISO 20294] International Organisation for Standardisation (ISO). Graphic technology — Quantification and communication for calculating the carbon footprint of e-media. (2018). https://www.iso.org/standard/67559.html.

[ISO 25024] International Organisation for Standardisation (ISO). Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality. (2015). https://www.iso.org/standard/35749.html.

[ISO 27001] The international standard (ISO/IEC) 27001 Information technology - Security techniques - Information security management systems.

[WP 29] Article 29 Data Protection Working Party. Guidelines on Consent under Regulation 2016/679. 10 April 2018. https://ec.europa.eu/newsroom/article29/items/623051.

[PIWIK PRO] Piwik PRO. Website: https://piwik.pro/blog/benefits-data-pseudonymization-anonymization-gdpr/

[MEHTA] Mehta Brijesha, Rao Udai Pratapb, Gupta Ruchikac, Conti Maurod. "Towards Privacy Preserving Unstructured Big Data Publishing." IOS Press Content Library. (2019): 3471 – 3482.

[MOTWANI] Motwani Rajeev, Nabar Shubha U. "Anonymizing Unstructured Data". Theory.stanford.edu. (2021)

[TERROVITIS] Terrovitis Manolis, Mamoulis Nikos, Kalnis Panos. "Anonymity in Unstructured Data". ResearchGate. (2008): 115-125

8 ANNEX 1

Data Management Plan Questionnaire - Version 1

Instructions

Deadline: May 21, 2019

For *each* data category/data type (these two terms are used as synonymous), you plan to generate, collect and/or process, please provide a **separate answer** to the following questions. For example, if you are going to process medical diagnostic data, and to generate software, you are managing two different data categories and you need to answer the below questions **for both data categories!**

We have already identified three data categories/data types are:

- Software code;
- Clinical/Health related Patient Data (such as Symptoms, Diagnoses, Laboratory Tests, and Genetic/Biometric Data);
- Other Patient Information (such as Patient Administrative Information – name, contact information, etc., Cultural/Religious/Ethnic Information, and Billing/Financial Information, mother tongue of patient).

If you are collecting/processing or generating additional categories of data, we encourage you to add them and answer the below questions also with respect to those additional categories.

The questions concern data collected and processed during the lifetime of the Project **only**. This questionnaire does not address the management of data once the envisioned applications are on the market. However, please note the questions address ***all* data types or categories and *not* just personal data** collected or generated during the Project.

If you are uncertain how to answer a question, please ask us:[6] dp-helpdesk.id@interopehrate.eu

Your answers to this Questionnaire will be seen in an annex to the Data Management Plan deliverable.

Please provide your answers in a different colour or in "revision mode" so that they are legible.

1. Data Types and Storage: The following questions are intended to understand what types of data will be generated and/or used in this project.

- What type of data will you **produce or generate** during the Project?
- What type of data will you **collect** during the Project?[7]
- How will you collect the data? In what formats?

[6] If you are not sure what information to include you may also take a look at some examples provided by other research institutions: https://ndownloader.figshare.com/files/9232003 Or https://depositonce.tu-berlin.de/bitstream/11303/8035/3/dmp_guidance_horizon2020_v1-0.pdf

[7] Deliverables are not included in this question.

- How will you trace the collected data?
  - *For example, how do you trace the provenance of the data collected? Alternatively, other metadata you maintain about the collected data.*

- Will the process of data generation or production be reproducible? What would happen if collected data got lost or became unusable later?

- How much data will it be, and at what growth rate? How often will it change?

- Are there tools or software needed to create/process/visualize the data?

- Will you use pre-existing data? From where?

- Storage and backup strategy?

2. Data Organization, Documentation and Metadata: The following questions are intended to understand the plan for organizing, documenting, and using descriptive metadata to assure quality control and reproducibility of these data.
Answer to the following questions only WRT the portion of data that you will publish (i.e. make available to people external to the project).

- What standards will be used for documentation and metadata (*e.g.,* Digital Object Identifiers)?
  - No standards?

- Do you use any best practices/guidelines for managing the data to publish (i.e., make available to third parties)?

- Do you use any tool for checking that the data are well formatted?

- What directory and file naming convention will be used?

- What project and data identifiers will be assigned?

- Is there a community standard for metadata sharing/integration?

- Can any metadata be created automatically?

3. Data Access and Intellectual Property
The following questions aim to identify any data access and ownership concern.

- What are the major risks to data security?

- What steps will be taken to protect privacy, security, confidentiality, intellectual property or other rights?

- Have you prepared a formal risk assessment addressing each of the major risks to data security and potential solutions?

- Does your data have any access concerns? Describe the process someone would take to access your data.

- Who checks the correct execution of the access process (e.g., PI, student, lab, University, funder)?
- What procedures have you developed for the safe transfer of personal or sensitive data?
- Any special privacy or security requirements (e.g., personal data, high-security data)?
- Any embargo periods to uphold?
- Have you implemented or outlined any procedures to follow in the case of a data breach?

4. Data Sharing and Reuse

The following questions are intended to clarify how the collected data will be released for sharing. *Answer to the following questions only WRT the portion of data that you will publish (i.e. make available to people external to the project)*

- If you allow others to reuse your data, how will the data be discovered and shared? List the categories of data that will be made re-usable or openly accessible.
- If so, how will you organize/label the data so that researchers may easily isolate fields of interest in their study (e.g., all women over 50 with hypertension)?
- Any sharing requirements? (e.g., funder data sharing policies often require that the digital data be released in machine-readable formats that supplement journal articles and presentations)
- Audience for reuse? Who will use it now? Who will use it later?
- Any restrictions on who can re-use the data and for what purpose?
- When will I publish it and where?

5. Data Preservation and Archiving

The following questions are intended to clarify how the collected data will be preserved and archived.

- How will the data be archived for preservation and long-term access?
- How long should it be retained (e.g., 3-5 years, 10-20 years, permanently)?
- What file formats? Are they long-lived?
- Are there data archives that are appropriate for your data (subject-based? Or institutional)?
- Who will maintain the data for the long-term?
- Who decides what data or what categories of data will be kept and for how long?
- The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. What protocol(s) will you put in place to ensure you delete personal data that is no longer required to be stored?

6. Ethical Aspects

- What types of personal data do you intend to collect, generate or process?
- What types of sensitive data do you intend to collect, generate or process?
- Will any of the data subjects be children or vulnerable people?
- Will you be collecting personal or sensitive data from people who have not given their explicit consent to participate in the Project?
- If you collected personal data, as defined by the GDPR, which of the six Art. 6.1 bases will you rely on for the processing of each category of personal data?
    - http://www.privacy-regulation.eu/en/article-6-lawfulness-of-processing-GDPR.htm
- If you collected sensitive data, as defined by the GDPR, which of the ten Art. 9 bases will you rely on for the processing of each category of sensitive data?
    - http://www.privacy-regulation.eu/en/article-9-processing-of-special-categories-of-personal-data-GDPR.htm
- Have you already gained consent for data preservation and sharing from any data subject(s)?
- How will you protect the identity of Project participants?
- Will you engage in large scale or big data processing?
- Will any entity (including any service provider) outside of the E.U. have access to personal or sensitive data?
    - If yes, who?
    - For what purpose?
    - Where is each of these entities located?

7. Pilots

Only for Pilot Leaders:
Who (or which entity or entities) will be responsible for determining what data is produced/generated for your Pilot?

Definitions and Other Reference Material

Personal Data means any information relating to an identified or identifiable natural person.

Special Category of Data is a subset of personal data, which includes data concerning health within the meaning of Art. 4 No. 15 of the GDPR, genetic data within the meaning of Art. 4 No. 13 and biometric data (e.g., fingerprints, facial images) if processed for the purpose of uniquely identifying a natural person (Art. 9).

Anonymization of data refers to the processing of personal data so as to irreversibly prevent identification of the natural person to whom the personal data is linked. For further information about anonymization see Art. 29 Working Party, Opinion 5/2014 on anonymization Techniques, WP 216 (2014), at 3.

Pseudonymisation of data refers to the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information (and

provided such additional information is kept separately). For a more detailed definition of pseudonymisation please refer to GDPR Art. 4 (5).

Processing means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available alignment or combination, restriction, erasure or destruction

Children mean any person below the age of 18.

Data Management Plan Questionnaire - Version 2

Instructions

The questions concern data collected and processed in the project as well as the management of data once the envisioned application is on the market. Please note the questions address all data types or categories and not just personal data collected or generated during the Project.

Your answers to this Questionnaire will be seen in an annex to the next version of the Data Management Plan deliverable.

If the answer is “none” or “not applicable” please state and provide a brief explanation of why there are no changes or why the question does not apply to you.

Please direct any questions to: iehr-helpdesk.id@univie.ac.at

Please submit the questionnaire by 1 September 2020.

1. General Questions:
    - The following answers are on behalf of? (e.g.: University of Vienna).
    - Please describe any significant changes to the categories of data you are processing/generating since the last DMP questionnaire (October 2019).
    - Please describe any significant changes to the way you process and/or generate data since the last DMP questionnaire.
    - Please describe any significant changes to data flows in which you are involved since the last DMP questionnaire.
    - Has your Data Protection Officer been involved in the setup of project-specific processing? Please provide the name and contact details of DPO.

2. Data Storage:
    - Since the previous DMP questionnaire (October 2019) what types of categories of data have you processed or generated?
    - What is the size of each category of data you are processing and/or generating and at what growth rate?
    - What storage and backup strategy has been adopted thus far, please indicate any relevant updates?

3. Data Security:
    - At this stage of the project, what are the major risks to data security with respect to the data categories you are processing and/or generating? And how did you assess these risks?
    - Please explain any risk mitigation measures you have taken and any relevant updates?
    - Please explain the privacy or security requirements you are using and/or observing?
    - Please explain what anonymization measures you are taking with respect to each of the data types you are processing.
    - Please explain whether there are any security or data minimization measures implemented with respect to DICOM images, movies or free text (especially in the case where this data represents patient information).
    - What procedure(s) do you have in place or will you adopt in the case of a data breach?

4. Data Access:
    - Do you plan to share the data you are processing and/or generating with any other partner (and/or any third party)?
    - If yes, have there been any changes to either the way you share data or the actual data you are sharing with others since the previous DMP Questionnaire?
    - Have any data access concerns been identified, if so which? (e.g.: unauthorised access)
    - Please describe the process implemented for data access?
    - Who checks the correct execution of the access process (e.g., PI, student, lab, University, funder)?
    - What procedures have you developed for the safe transfer of personal or sensitive data?

5. Data Sharing and Re-Use:
    - What tools or platforms are being used to facilitate data sharing between consortium partners? (e.g., GitLab, Google Drive, etc.)
    - Do you use third party services (including storage services) to assist with data processing?
    - If you answered "Yes" to the above question, please explain what third party services you use and for what purpose(s).

Questions relating to Pilot Sites:

1. Data Collection at Pilot Sites:
    - What data sets will be collected from volunteer patients and in what formats? (e.g., contact data, electronic health records in word format - please specify what type of information - prescription info? disease or surgery history? location information?, DICOM images, in an excel file, etc.).
    - How many volunteer patients will be involved in your pilot?
    - Are there tools or software needed to process and visualise the data you plan to collect as part of the pilots?
    - Will you always require informed consent before collecting volunteer patients' data?

- How will you ensure compliance with GDPR Articles 13 through 18 and 20 through 22, and Article 34 (regarding information required to be provided to patients, rights of access, rectification, erasure, etc.)?

2. Data Storage at Pilot Sites
    - Where and how will you store volunteer patients' data? (e.g., S-EHR Cloud)
    - How long will you store volunteer patients' data?
    - Please explain the backup strategies to be adopted.
    - What protocol(s) will you put in place to ensure that personal data is not stored for longer than is required?

3. Data Security at Pilot Sites
    - What are the major risks to data security during the pilot phase? And how did you assess those risks?
    - Please explain the risk mitigation measures you have or plan to undertake?
    - Please explain the privacy or security measures being applied to secure volunteer patients' data? (e.g., encryption, anonymization of data).
    - Further explain the use of anonymization or pseudonymisation techniques at the pilot sites and the implementation of these techniques in the case of DICOM images, movies and/or free text.
    - What procedure have you adopted or will you adopt to address the possibility of a data breach?

4. Data Access at Pilot Sites
    - Will you share any personal data or health related data with any person or entity outside of your hospital/organization? If so, whom?
    - Have any data access concerns been identified, if so which? (e.g., unauthorised access)
    - Please describe the process implemented for data access and who will be granted access?
    - Who checks the correct execution of the access process?
    - Will the patient have access to the data they provide (for purposes of, for example, correcting errors)?
    - What procedures have you developed for the safe transfer of personal or sensitive data?

5. Data Sharing and Re-Use at Pilot Sites
    - Will data from pilot sites be reused? If so, who is the audience for reuse? Who will use it now? Who will use it later?
    - Any restrictions on who can re-use the data and for what purpose?
    - How will the data be shared? List the categories of data that will be made re-usable or openly accessible.
    - Any sharing requirements?
    - Will any data or results from the pilot sites be published and where?

6. Data Preservation and Archiving at Pilot Sites
    - Will the data collected at pilot sites be archived in your organisation's systems for preservation and long-term access? If so, how will this be achieved?
    - If so, for what purposes will such data be maintained after the end of the project?
    - What are the file formats? Are they long-lived?
    - How long will it be retained?
    - How will such data be managed at the end of the project?
    - Who will maintain the data for the long term?

9 ANNEX 2

Date: _____

Form: Declaration of Compliance

As the Representative for _____(the "Consortium Partner"), I _____ have been provided with information relating to the data processing operations that may be carried out by the Consortium Partner and its members in order to fulfil the obligations within the InteropEHRate project.

I have reviewed the text of the General Data Protection Regulation (GDPR), including provisions in the GDPR that grant Member States legislative competence, and I am familiar with the GDPR implementation act of the country in which this Consortium Partner resides. I confirm that this Consortium Partner is in compliance and will act in accordance with the GDPR and all national GDPR implementation acts that are applicable to the InteropEHRate project.

I have checked if special derogations pertaining to the rights of data subjects or the processing of biometric and/or health data have been established under the national legislation of the country where the research takes place. The following national derogations and national legislations apply in this project:

[PLEASE SPECIFY BELOW, THE RELEVANT COUNTRY AND THE NATIONAL LEGISLATION APPLICABLE]

I can confirm that the Consortium Partner and its members who are involved in the InteropEHRate project are fully aware of their obligations in this regard.

Name: _____

Signature: _____

Contact Details: _____